

Summer 8-1-2011

Utility of the Inventory of Callous-Unemotional Traits in Adolescent Offenders and Non-Offenders: An Item Response Theory Analysis

Lisa Laurence Ansel
University of Southern Mississippi

Follow this and additional works at: <https://aquila.usm.edu/dissertations>

Recommended Citation

Ansel, Lisa Laurence, "Utility of the Inventory of Callous-Unemotional Traits in Adolescent Offenders and Non-Offenders: An Item Response Theory Analysis" (2011). *Dissertations*. 562.
<https://aquila.usm.edu/dissertations/562>

This Dissertation is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Dissertations by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.

The University of Southern Mississippi

UTILITY OF THE INVENTORY OF CALLOUS-UNEMOTIONAL TRAITS IN
ADOLESCENT OFFENDERS AND NON-OFFENDERS:
AN ITEM RESPONSE THEORY ANALYSIS

by

Lisa Laurence Ansel

Abstract of a Dissertation
Submitted to the Graduate School
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

August 2011

ABSTRACT

UTILITY OF THE INVENTORY OF CALLOUS-UNEMOTIONAL TRAITS IN ADOLESCENT OFFENDERS AND NON-OFFENDERS: AN ITEM RESPONSE THEORY ANALYSIS

by Lisa Laurence Ansel

August 2011

The current study utilized Item Response Theory (IRT) analyses to examine the item functioning of the Inventory of Callous-Unemotional Traits (ICU; Frick, 2003) in the assessment of psychopathy-linked traits in adolescents. Self-report psychopathy measures, such as the ICU, have become increasingly common for use with adolescents. However, questions remain regarding their reliability and utility for accurately assessing these traits (Poythress, Dembo, Wareham, & Greenbaum, 2006). IRT analyses offer unique methods of investigating test and item functioning in regards to the underlying trait an inventory purportedly assesses. The current study examined and compared the item functioning of the ICU for the assessment of psychopathy-linked characteristics, particularly callous-unemotional (CU) traits, in a sample of adolescent offenders as well as a non-offending adolescent sample. Based on the current results, ICU items 11, 15, 16, and 23 were the most discriminating items in the non-offender sample. Items 5, 15, 17, and 23 were the most discriminating in the offender sample. Items 11 and 17 were found to be the most difficult and discriminating items for the non-offender and offender samples, respectively. The results of the current study further indicated that eight of the ICU items functioned differently between the two samples in either discriminating ability or item difficulty. These results support the notion that differential item functioning exists

between the current sample of adolescent offenders and non-offenders. Overall, the ICU appears to estimate CU traits more reliably in this sample of non-offenders compared to the adolescent offender sample.

COPYRIGHT BY
LISA LAURENCE ANSEL
2011

The University of Southern Mississippi

UTILITY OF THE INVENTORY OF CALLOUS-UNEMOTIONAL TRAITS IN
ADOLESCENT OFFENDERS AND NON-OFFENDERS:
AN ITEM RESPONSE THEORY ANALYSIS

by

Lisa Laurence Ansel

A Dissertation
Submitted to the Graduate School
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

Christopher Barry, Ph.D.
Director

Randolph Arnau, Ph.D.

David Marcus, Ph.D.

Sterett Mercer, Ph.D.

Susan A. Siltanen, Ph.D.
Dean of the Graduate School

August 2011

ACKNOWLEDGMENTS

The writer would like to thank, first and foremost, Dr. Christopher Barry. Over the past several years, Dr. Barry has given unwavering guidance and support through the hardships and triumphs of graduate school. Dr. Barry has willingly served as both my authoritative leader and enthusiastic cheerleader. I am deeply thankful for his endless knowledge and direction, especially through the hurdles of the dissertation process. Without Dr. Barry, I would not have reached the finish line!

I would also like to thank each of the remaining committee members, Dr. Randolph Arnau, Dr. David Marcus, and Dr. Sterett Mercer, for their willingness to share their incredible expertise and precious time in order to guide and improve my dissertation project. I am grateful for the valuable supervision I received from the brilliant members of my committee. In addition, I would like to give very special thanks to Dr. Eva Kimonis and Dr. Monica Marsee. Due to their work and dedication to research, what was once just an idea quickly transformed into a doctoral dissertation project. It would not have been possible without their contributions!

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	iv
LIST OF TABLES.....	vi
LIST OF ILLUSTRATIONS.....	vii
CHAPTER	
I. INTRODUCTION.....	1
II. REVIEW OF RELATED LITERATURE.....	3
Callous-Unemotional (CU) Traits	
Inventory of Callous-Unemotional Traits (ICU; Frick, 2003)	
Item Response Theory	
Summary and Current Study's Hypotheses	
III. METHODOLOGY.....	24
Participants	
Materials	
Procedure	
IV. ANALYSIS OF DATA.....	28
Assumptions of IRT	
IRT Model	
Descriptive Statistics	
Factor Analyses	
IRT Analyses	
V. DISCUSSION.....	48
IRT Evaluations of Item and Test Functioning	
CU Traits as Assessed by the ICU	
Limitations and Directions for Future Research	
Conclusion	
APPENDIX	62
REFERENCES.....	63

LIST OF TABLES

Table

1.	Demographic and Descriptive Statistics.....	31
2.	Confirmatory Factor Analyses (CFA) Results.....	32
3.	Exploratory Factor Analysis (EFA) Results.....	34
4.	Item Parameter Estimates for Independent Samples.....	36
5.	Item Parameters for Offender Sample Alone and Combined Offender and Non-Offender Model With Common Metric	45

LIST OF ILLUSTRATIONS

Figure

1.	Item Characteristic Curves for ICU Items in Offender Sample.....	40
2.	Item Characteristic Curves for ICU Items in Non-Offender Sample.....	41
3.	Total Test Information for Offender Sample.....	43
4.	Total Test Information for Non-Offender Sample.....	44

CHAPTER I

INTRODUCTION

Psychopathy refers to a constellation of personality and behavioral characteristics associated with maladaptive functioning including severe antisocial behavior. Individuals exhibiting elevated levels of psychopathy-linked characteristics are at a greater risk for criminal reoffending, committing violent offenses, and being nonresponsive to treatment and intervention programs (Loney, Taylor, Butler, & Iacono, 2007). In fact, psychopathy in adults is considered to be one of the primary indicators of increased risk for continuing criminal behavior (Douglas, Vincent, & Edens, 2006). Results of a meta-analysis including 18 studies suggest that psychopathy is a useful predictor of both general (nonviolent) and violent reoffending (Salekin, Rogers, & Sewell, 1996). Furthermore, psychopathy theoretically represents a collection of relatively stable traits, beginning in childhood and continuing across the lifespan (Forth, Hart, & Hare, 1990). One study revealed that elevated psychopathy scores in early adolescence were predictive of recidivism over a three- to four-year follow-up period (Salekin, 2008), suggesting that personality and behavioral characteristics indicative of psychopathy are not likely to desist during this critical developmental period.

Due to the relations between psychopathic characteristics, maladaptive interpersonal functioning, and criminal behavior, accurate assessment of psychopathic traits is necessary to aid in prevention efforts and intervention development. Of particular interest to the current study is the ability of a specific measure of psychopathy-linked characteristics (i.e., Inventory of Callous-Unemotional Traits), as well as individual items

of the measure, to discriminate between a sample of adolescents with a positive history of illegal offenses and a non-offending community sample of adolescents.

CHAPTER II

REVIEW OF RELATED LITERATURE

Previous research distinguishes between two primary components of psychopathy commonly referred to as Factor 1 and Factor 2. Factor 1 includes personality traits, such as callousness, lack of remorse and guilt, manipulateness, flat affect, and selfishness (Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007; Vitale, Smith, Brinkley, & Newman, 2002). Factor 2 focuses on behavioral features including irresponsibility, impulsivity, and risk-taking. This 2-factor conceptualization has guided the development of assessment instruments, such as the Psychopathy Checklist (PCL; Hare, 1980) and later counterparts, including its revision (PCL-R; Hare, 1991) and the Psychopathy Checklist-Youth Version (PCL-YV; Forth, Kosson, & Hare, 2003). These instruments have largely dominated the evaluation and research of psychopathy-linked traits and are considered by many to be the “gold-standard” for the assessment of psychopathy (Jones, Cauffman, Miller, & Mulvey, 2006, p. 34). However, as interest in the field of psychopathy has grown, so has the development of additional assessment approaches aimed at identifying levels of psychopathy-linked traits in adolescents in general and juvenile offenders, specifically.

Self-report instruments of psychopathy have been increasingly utilized by researchers and clinicians. These measures provide a more efficient method of screening of psychopathy-linked characteristics—in contrast to the interview format of the PCL and its revisions which rely on clinician ratings and an extensive file review—in that they require little training for the interviewer and are significantly less time-consuming for both the administrator and the adolescent. Furthermore, they allow information to be

obtained from the individual regarding personality and affective characteristics that may not be easily or accurately identified by an outside observer. Specifically, many of the affective characteristics central to the construct of psychopathy, such as callousness, a lack of remorse, and a lack of care for others, cannot be evaluated solely through observations and reports of others.

Despite the many advantages of using self-report psychopathy measures, research is relatively limited regarding their psychometric properties. Salekin (2008) described research results indicating no differences between the self-report Antisocial Process Screening Device (APSD; Frick & Hare, 2001) and the interview-based PCL:YV regarding predictions of general recidivism. Correlational analyses also revealed a significant moderate correlation between the APSD and PCL:YV total scores (i.e., $r = .37, p < .01$). Furthermore, the APSD demonstrated an acceptable internal consistency ($\alpha = .71$) for this sample of adolescents. In short, adolescent self-report via a 20-item inventory was as useful as clinician-based interviewing for predicting recidivism (Salekin, 2008).

Poythress and colleagues (2006) examined the internal consistency of the APSD and the Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Levander, 2002) in a sample of justice-involved adolescents. For this sample, the APSD Narcissism and Impulsivity scales demonstrated moderate reliability ($\alpha = .61$ and $.57$, respectively). The APSD Callous-Unemotional (CU) scale displayed an even lower alpha level of $.45$. Similar results were obtained for the YPI in this sample with the Callousness subscale demonstrating the lowest internal consistency ($\alpha = .36$) of all the remaining YPI subscales (Poythress, Dembo, Wareham, & Greenbaum, 2006). These results indicate

that current self-report psychopathy measures may not be adequately assessing traits associated with callousness in an adolescent sample. In another study (Falkenbach, Poythress, & Heide, 2003), both the self-report APSD and the Child Psychopathy Scale (CPS; Lynam, 1997) demonstrated predictive validity for predicting treatment compliance and recidivism in a sample of adolescent offenders referred to a diversion program. These results provide some initial evidence that relatively short self-report measures of psychopathy-linked traits in adolescents have validity in the assessment of psychopathy as well as clinical utility in the prediction of future antisocial behavior (Falkenbach et al., 2003; Salekin, 2008). However, further research is clearly needed regarding the psychometric properties and utility of such measures (Poythress et al., 2006).

Callous-Unemotional (CU) Traits

Previous research suggests that individuals possessing high levels of psychopathy-linked characteristics can be distinguished from individuals exhibiting antisocial behavior in general based on the presence of CU traits (Frick, 2003). CU traits are thought to comprise at least part of Factor 1 of psychopathy described above. The combination of high levels of conduct problems and self-reported CU traits has been linked to more severe and varied problem behaviors in youth as compared to the presence of conduct problems alone (Frick, Cornell, Barry, Bodin, & Dane, 2003). Children with both CU traits and conduct problems were also more likely than children without CU traits to have high levels of aggression and self-reported delinquency. In another study, children and adolescents exhibiting conduct problems without elevated CU traits had lower frequencies and types of behavior problems, fewer police contacts, and a lower

likelihood of having a parent with Antisocial Personality Disorder than youth with comparable levels of behavioral problems but higher levels of CU traits (Christian, Frick, Hill, Tyler, & Frazer, 1997). Based on accumulating evidence, elevated levels of CU traits in adolescents have been considered one of the primary indicators of adolescents who are likely to continue antisocial behaviors into adulthood (Falkenbach et al., 2003). The risk of later behavioral problems associated with CU traits is noteworthy in that CU traits are largely affective and interpersonal rather than denoting a set of behaviors that places an individual at-risk for later, similar behaviors.

However, some researchers caution that the presence of psychopathy-linked traits in adolescence does not automatically indicate that those individuals will engage in a persistent pattern of maladaptive or antisocial behavior. According to Moffitt (2003), delinquent behavior in adolescents may be normative and a natural part of adolescent development in that these types of behaviors may serve as a means of gaining approval from peers and seeking independence from parents or authority figures. More specifically, some of the items included on psychopathy measures, such as impulsivity and sensation-seeking, may represent behavioral or personality features that are consistent with typical adolescent functioning and will become less pronounced as the adolescent transitions into adulthood (Edens, Skeem, Cruise, & Cauffman, 2001). Edens and colleagues (2001) further suggest that due to personality and emotional systems that are not fully developed, adolescents may appear to possess some of the primary features of psychopathy (e.g., shallow affect) to outside observers. Therefore, despite its potential benefits, using scores from psychopathy assessment measures to inform decisions

regarding legal, clinical, or placement issues for adolescents may be misleading and could potentially lead to “false positives.”

Nevertheless, it appears that accurate distinctions between normative presentations of CU traits during adolescence versus levels of CU traits indicating an increased likelihood of antisocial behavior are still relevant for predicting problematic behavior and subsequent attempts at prevention/intervention. Clinicians and legal professionals should, however, be aware of the limitations of these assessment measures when considering the best interests of an adolescent. Simply using total scores from existing assessment instruments may inflate the presence of psychopathy, particularly for adolescents. Thus, also considering results at the item-level, although potentially less efficient, may be a more sound way to arrive at a better understanding of the connection between specific facets of psychopathy and severe behavioral problems as well as to improve the decisions that are made based on assessment results. The current study investigated which, if any, of the individual CU traits assessed by a self-report measure yielded particularly useful and discriminating information in two adolescent samples (i.e., offender versus non-offender).

Inventory of Callous-Unemotional Traits (ICU; Frick, 2003)

The ICU, developed by Frick (2003), is one of the most recently developed self-report measures assessing the presence of CU traits in youth. In contrast to the widely used APSD, which includes items assessing both personality and behavioral features, the ICU provides a more specific focus on CU traits. Though the construct of psychopathy includes maladaptive behavioral features (e.g., impulsivity), the author of the ICU intended to provide a comprehensive assessment of affective traits, particularly

callousness, lack of care for others (i.e., “uncaring”), and unemotionality (Essau, Sasagawa, & Frick, 2006) which are considered to be distinguishing indicators of psychopathy (Cleckley, 1976; Hare, 1998). The 24 items comprising the ICU are rated by the individual using a four-point Likert scale with responses ranging from 0 (*not at all true*) to 3 (*definitely true*). The first examination of the ICU’s psychometric properties was conducted by Essau and colleagues (2006) in a sample of German adolescents. The results of statistical analyses reported by Essau and colleagues (2006) suggest that the items of the ICU can be grouped into three factors or dimensions (i.e., Callousness, Uncaring, Unemotional) which are combined to obtain the total score. Examples of Callousness items include “*I do not care who I hurt to get what I want*”, “*I do not feel remorseful when I do something wrong*”, and “*The feelings of others are unimportant to me*” (Frick, 2003). Examples of items assessing Uncaring are “*I try not to hurt other’s feelings*” and “*I apologize to persons I hurt*” which are both reverse-scored. Items such as “*I express my feelings openly*” and “*I am very expressive and emotional*” are included to assess unemotionality. In a separate study, the factor structure of the ICU was examined in a sample of juvenile offenders detained in various facilities in the Southeastern United States (Kimonis, Frick, Skeem et al., 2008). The resulting factor structure was largely consistent with the factor structure presented by Essau and colleagues (2006) in such that three separate subfactors were found to be related to the total ICU scale. However, it was necessary to delete two items (i.e., ICU items 2 and 10) to obtain adequate model fit based on poor item-total correlations (Kimonis, Frick, Skeem et al., 2008).

Although the ICU is one of the most recently developed self-report measures of psychopathy-linked traits in adolescents, research regarding the ICU's reliability and validity in a variety of samples (e.g., community, at-risk) is becoming increasingly available. Previous research examining the relations between the ICU and externalizing behaviors in adolescents has provided evidence for the validity of the ICU in that the total ICU score, as well as the Callousness and Uncaring subscales, were significantly correlated with externalizing behaviors, such as aggression and delinquency (Essau et al., 2006; Kimonis, Frick, Skeem et al., 2008). ICU total scores were also found to be significantly correlated with lower emotional reactivity (Kimonis, Frick, Skeem et al., 2008). However, the Unemotional subscale was negatively related to externalizing behaviors in one sample of adolescents (Essau et al., 2006) and not strongly related to aggression or delinquency in another sample (Kimonis, Frick, Skeem et al., 2008).

Despite the growing literature regarding the ICU, there is no known study to date examining the psychometric properties of the individual items comprising the ICU in terms of each item's ability to differentially discriminate between CU trait levels in adolescents convicted of criminal or status offenses and non-offending adolescents from a community sample. The current study investigated whether items of the ICU provided differential information depending on an adolescent's status of offender or non-offender. Determining which items, and presumably which aspects of CU traits, provide the most behaviorally relevant information has important implications for understanding the developmental manifestation of psychopathy in adolescence, for improving the assessment of these characteristics, and in the development of prevention and intervention strategies.

Item Response Theory

Item Response Theory (IRT) is a test measurement model designed to address several of the shortcomings associated with Classical Test Theory (CTT) models. IRT is considered a latent-trait model because of the assumption that responses to items of a test or measurement instrument are the result of a specific underlying trait (e.g., ability or psychological construct) which is represented symbolically by theta (θ) in an IRT analysis (Zickar & Broadfoot, 2009). It is assumed that the underlying trait exists on a continuum ranging from low extremes to high extremes and that every respondent will fall somewhere along the continuum as opposed to either fitting or not fitting into a category. According to IRT, the level of the underlying trait explains or predicts the response of an individual on any given test item. Furthermore, the relation between an individual's response on a particular item, referred to as item performance, and the underlying trait is described by a mathematical function labeled an Item Response Function or an Item Characteristic Curve (ICC). Theoretically, the probability of an individual endorsing an item (e.g., providing a response that indicates higher levels of the construct) increases as the level of the underlying construct increases (Hambleton, Swaminathan, & Rogers, 1991).

One assumption of IRT is that the measurement instrument is unidimensional (Hambleton et al., 1991). Unidimensionality suggests that there is only one latent trait underlying item responses. For example, an individual's performance on a mathematics ability test is solely influenced by the unidimensional nature of the examinee's knowledge regarding mathematical rules and operations assessed by the items on the test. Despite this assumption, previous researchers have utilized IRT to analyze tests shown to

have more than one factor (Bolt, Hare, Vitale, & Newman, 2004; Cooke & Michie, 1997). For example, Cooke and Michie (1997) examined item and test functioning of the PCL-R by separately analyzing the PCL-R total score, Factor 1, Factor 2, and the individual items. The rationale behind these analyses stems from the conceptualization of the total test score as a representation of the overall construct of psychopathy which is composed of two correlated factors (Cooke & Michie, 1997). Therefore, due to the expected significant correlations between separate factors of a test, IRT analyses are still deemed appropriate. In this way, a test may be considered both unidimensional and multidimensional if the individual factors can be explained by a higher-order underlying trait (McDonald, 2000).

The second primary assumption of IRT is independence of scores or what is termed local independence. This assumption purports that the items comprising the measurement instrument are only related to all other items of the instrument because of the underlying trait. According to the assumption of local independence, significant correlations between pairs of test items would cease to exist if the effects of the underlying trait were partialled out (Zickar & Broadfoot, 2009). In other words, items from a particular test are considered to have local independence if the test items are statistically independent across all levels of the latent trait (Vincent, 2002). Refer to the Analyses section below for a description of procedures utilized to check for these assumptions.

In addition to analyzing the amount of information provided by individual test items, researchers are able to investigate entire assessment instruments even more specifically through IRT analyses. Statistical conclusions can be made regarding which

items can or should be deleted due to a lack of unique information about the construct provided by particular items. Analyzing the plot lines of the ICCs may reveal which items are functioning as maximum predictors for the construct or trait at a given trait level. For example, a steep slope of the ICC for a specific item at a given trait level suggests that as the level of θ increases past a certain point, the probability of endorsing that item increases. In other words, an individual possessing low levels of the underlying trait has a decreased probability of endorsing an item that is discriminating at the higher end of the continuum. For diagnostic purposes, an item with maximum discrimination would be able to distinguish between individuals with and without a particular disorder depending on the stated diagnostic cutoff (Cooke & Michie, 1997).

Furthermore, researchers have used IRT analyses to investigate the extent to which item bias exists on a given measure. When a test or test item is referred to as biased, it usually means that the item or test performs differently for different groups of individuals (Cooke & Michie, 1997). This information may be used to identify tests that produce different scores across groups due to undesired biases, such as race or gender bias. On the other hand, item bias may also be referred to as Differential Item Functioning (DIF) which, if it exists, may be beneficial depending on the purpose of the test. For example, DIF may occur if an individual item yields different scores for one group as compared to another group of individuals with comparable levels of the underlying trait. For the purposes of the current study, DIF was utilized as an indicator that certain items on the ICU discriminate differently between a sample of adolescent offenders as compared to adolescents from a community sample.

Depending on the nature of the measure under investigation, as well as the research questions the investigator seeks to address, there are several different IRT models that can be utilized which will determine the shape of the ICC. The most basic description of IRT models begins with a discussion of the parameters chosen by researchers conducting an IRT analysis. First, the b parameter refers to the *item difficulty* parameter. Item difficulty is conceptualized as the location of θ where a particular item is most discriminating or provides the most information (Zickar & Broadfoot, 2009). Using the construct of psychopathy as an example, the probability of endorsing an item with low difficulty would be the same for individuals with low levels of psychopathy-linked characteristics as well as those individuals possessing high levels of psychopathic traits. Conversely, only those individuals at the highest levels of psychopathy, or θ , would endorse items with high difficulty.

The a parameter, named the *item discrimination* parameter, provides information regarding an item's ability to differentiate between respondents with varying levels of θ (Hambleton et al., 1991). An item's discriminating ability is demonstrated as the slope of the ICC at the point of inflection. The steeper the slope of the ICC for an individual item, the higher the discriminating power of that item. In other words, an item with high discriminating power enables the researcher to make distinctions between individuals with different levels of the underlying trait. If an item's ICC slope is flat, that item is considered to have low discrimination and is not useful for distinguishing between respondents with varying trait levels. Furthermore, the amount of item discrimination will be influenced by the range of the respondents' underlying trait levels. Specifically, one item may have high discrimination at lower levels of θ and low discriminating ability

when considered at the highest levels of the trait (Hambleton et al., 1991; Zickar & Broadfoot, 2009). In IRT analyses, the combination of the a and b parameters is referred to as a two-parameter item response model which is one of the most commonly used models in IRT analyses (Hambleton et al., 1991).¹

One of the specific pieces of information that may be obtained following an IRT analysis is the Item Information Function. Item Information Functions can be used to describe items, to select specific items to include on a test based on the amount of information provided by the item, and to compare tests with each other (Hambleton et al., 1991). Information refers to how much an individual item contributes to the total amount of information necessary to make predictions or estimations. In other words, item information is considered to be an estimate of the reliability of the item at different levels of the underlying construct (Cooke & Michie, 1997). Similar to item discrimination, an item may provide substantial information at low levels of θ , though it does not offer much information regarding higher levels of the trait. Information is directly influenced by an item's discrimination and difficulty. When graphed, the height of the item information function is congruent with the item discrimination at a given level of θ . Furthermore, the point at which the item information function peaks will be close to that item's difficulty parameter (Zickar & Broadfoot, 2009). Summing all of the item information functions provides the researcher with the Test Information Function. The test information function describes how useful the test is for accurately estimating respondents' trait levels at each level of θ . Ideally, tests designed to assess the presence, or extent, of a trait for a group of respondents should produce maximized information

across a wide range of the trait (i.e., for individuals endorsing low and high levels of the trait; Cooke & Michie, 1997).

Application of IRT to Psychopathy Assessment Measures

Cooke and Michie (1997) conducted one of the first studies using IRT to investigate test and item functioning of a psychopathy measure. In this study, the researchers investigated the PCL-R (Hare, 1991) in terms of its ability to provide information regarding a diagnosis of psychopathy in an adult sample as well as to measure trait strength. Cooke and Michie based analyses on the assumption that the total PCL-R score represents an estimate of the construct of psychopathy, whereas the two factor scores (i.e., Factor 1 and Factor 2) represent separate but related dimensions of a psychopathic personality disorder. One purpose of this study was to assess the usefulness of individual PCL-R items for guiding diagnostic decisions. Previous use of the PCL-R in clinical settings has relied on the total score, giving little attention to which psychopathy-linked characteristics, as assessed by individual items, were endorsed. This approach assumes that ratings on each item are equally representative of the underlying level of psychopathy (Cooke & Michie, 1997). However, it has been suggested that the number of items endorsed is not necessarily related to an underlying trait in a linear fashion (Nunnally & Bernstein, 1994). Certain features assessed by a test may be more important, or more discriminating, at higher levels of the underlying trait or construct.

Cooke and Michie used data from ten samples (combined to create two Canadian and two American samples) of adults in forensic or psychiatric settings. Because there are three possible responses for the PCL-R items, ranging from 0 (*does not apply*) to 2 (*definitely applies*), Samejima's Graded Response Model (GRM; Samejima, 1969)

consisting of 2-parameter logistic functions (i.e., a and b parameters) was utilized. Using Multilog 6 (Thissen, 1991), the researchers assessed the generalized likelihood ratio testing (GLRT) for the four samples. The GLRT was used to provide an indicator of whether the IRT parameters were essentially equal across the four samples. Results indicated no significant DIF across these samples, suggesting that the four samples could be described using the same parameters. Upon investigation of the difficulty parameters (b_1 —the probability of a 0 response, and b_2 —the probability of a 2 response), the researchers concluded that the test items provided high discrimination across various levels of the underlying trait of psychopathy, providing support for use of the PCL-R as a measure of trait strength (Cooke & Michie, 1997).

At the item level, results indicated that items assessing affective features such as callousness, shallow affect, and lack of remorse were particularly discriminating as evidenced by large item discrimination, a , parameters. The least discriminating items included those assessing juvenile delinquency, many short-term relationships, and promiscuous sexual behavior. Overall, the item difficulty parameters indicated that Factor 1 items had higher threshold parameters than Factor 2 items (Cooke & Michie, 1997). These findings suggest that affective traits, such as CU traits, become increasingly important at higher levels of psychopathy. The authors concluded that the PCL-R is adequate for assessing psychopathic personality and that characteristics assessed by Factor 1 are more likely to occur at higher levels of the underlying psychopathy trait. The authors interpret these results as further evidence that Factor 1 items, including CU traits, are central to the construct of psychopathy and warrant significant attention (Cooke & Michie, 1997).

Vincent (2002) also utilized IRT analyses to investigate the assessment of psychopathy in adults and adolescents. Specifically, the researcher was interested in whether adolescent psychopathy represents a legitimate construct and used IRT to examine the extent to which age-related measurement bias existed on the PCL:YV. Previous research regarding the psychometric properties of the PCL scales suggests that adult psychopathy, as assessed by the PCL, represents a clear and stable syndrome (Forth, Brown, Hart, & Hare, 1996). In contrast, there has been a relative lack of evidence of a coherent, stable psychopathy syndrome in adolescents. Furthermore, Vincent (2002) sought to address the concern that little is known about the long-term stability of psychopathy-linked traits in adolescents and the possibility that the measurement and assessment of these traits in adolescents is not analogous to that of adults.

Based on the notion that many test items on adolescent psychopathy assessment measures may represent developmentally appropriate characteristics, Vincent (2002) employed psychometric techniques including an evaluation of the structural validity of the PCL:YV as well as its item and test functioning in relation to adult assessments using the PCL-R. Specific questions included whether psychopathy represents a coherent construct in adolescents, which items, or traits, are particularly meaningful in youth, and whether an age-related measurement bias exists for adolescents. The sample consisted of 269 male adolescents and 444 adult males. Consistent with the approach of Cooke and Michie (1997), Multilog 6 (Thissen, 1991) was used to assess item and test performance.

In addition to the 2-factor model of the PCL discussed previously, factor analyses have also yielded a 3-factor model for the PCL:YV (Kosson, Cyterski, Steuerwald, Neumann, & Walker-Matthews, 2002). Therefore, Vincent (2002) conducted factor

analyses to determine which model yielded the best fit for both 2- and 3-factor models as well as met the IRT assumption of unidimensionality in this sample. Factor analyses revealed the best fit for a 3-factor model in both the adult and adolescent samples which also met criteria for unidimensionality. The three factors were identified as *Arrogant and Deceitful Interpersonal Style*, *Deficient Affective Experience*, and *Impulsive and Irresponsible Behavioral Style*. Regarding age-related measurement bias, which was indicated by DIF, the results suggest that measurement bias did exist dependent on age for interpersonal and behavioral psychopathy-linked characteristics in such a way that adolescents needed to have higher levels of θ , compared to adults, in order to be rated as highly as adults on these features. However, there was no significant DIF regarding the affective features, such as remorselessness or callousness/lack of empathy. Furthermore, an item indicating callousness/lack of empathy was the most discriminating for both the adolescent and adult samples, as indicated by the largest a parameter. High discrimination suggested that individuals with low levels of psychopathy characteristics were unlikely to receive high scores on this item; therefore, high scores on this item appeared to be a reliable indicator of psychopathy in that sample (Vincent, 2002). Impulsivity, on the other hand, appeared to be the least discriminating item for that sample of adolescents. The author interpreted these results as potentially supporting evidence that the affective features of psychopathy, such as callousness and remorselessness, “may be both a *necessary* and a *sufficient* condition of this disorder” (Vincent, 2002, p. 81) in both adolescents and adults.

IRT analyses were also conducted, in a separate study, to determine the item and test functioning of the PCL:YV in a sample of detained adolescent females (Schrum &

Salekin, 2006). The authors investigated the appropriateness of the application of psychopathy to adolescents, particularly adolescent females. In addition to the concerns of applying this construct to adolescents highlighted previously, the authors were specifically interested in the manifestation of psychopathy in female adolescents. Schrum and Salekin (2006) suggested that an increased understanding of this construct in females is particularly important in light of the growing prevalence rates of female juvenile offenders as well as the relative lack of research surrounding psychopathy assessment in females.

In contrast to previously discussed studies, Schrum and Salekin (2006) utilized a two-factor model of the PCL:YV, which was subsequently divided into four facets, as opposed to the three-factor model used by Vincent (2002). The four facets were conceptualized as representing Interpersonal, Affective, Lifestyle, and Antisocial features of the higher-order psychopathy construct (Schrum & Salekin, 2006). The sample consisted of 123 female adolescents mandated to juvenile detention centers in the southern United States. Consistent with previously discussed IRT analyses, the authors utilized Multilog 6 (Thissen, 1991) and Samejima's GRM (Samejima, 1969) to analyze item and test functioning. Regarding discriminating ability of individual test items, those items assessing grandiosity, manipulativeness, and callousness/lack of empathy produced the largest a parameters signifying the most discriminating power for this sample of female adolescents. Items purported to assess poor anger control, shallow affect, and probation violations emerged as the least discriminating items for this sample (Schrum & Salekin, 2006). Analyses of parameter means at the factor level revealed that, consistent with previous research, Factor 1 was a better discriminator of psychopathy than Factor 2.

Furthermore, Factor 1 demonstrated the most discriminating power at the highest levels of psychopathy and the greatest amount of overall information. Among the four facets, the Interpersonal facet resulted in the largest a and b parameters indicating that this facet was the most discriminating and provided the most information regarding the underlying construct. Of the remaining three facets, the Affective facet offered the most information, followed by the Antisocial facet, with the Lifestyle facet providing the least amount of information (Schrum & Salekin, 2006).

Based on these results, the authors concluded that items assessing callousness, grandiose sense of self-worth, and manipulateness are useful for assessing the underlying construct of psychopathy in this sample of adolescent females because they are unlikely to be endorsed at low levels of the trait. Conversely, there is a high likelihood for these items to be endorsed at the highest levels of psychopathy. Overall, these results support the extension of the psychopathy construct to adolescent females. However, due to the lack of information regarding the level of the underlying trait provided by the behavioral features of psychopathy (i.e., the Lifestyle and Antisocial facets) for this sample, the authors caution against relying on high scores on behavioral items when assessing for the presence of psychopathy. Consistent with previously discussed research studies involving adolescent males and adult samples, the authors conclude that high scores on behavioral items may artificially inflate total psychopathy scores for adolescent females (Schrum & Salekin, 2006).

Summary and Current Study's Hypotheses

The application of Item Response Theory as a psychometric tool for investigating psychopathy assessment instruments has only recently begun to hold a place in the

research literature. As evidenced by the previous review of relevant research, the majority of these IRT studies have focused on some version of the PCL (e.g., PCL-R, PCL:YV), which is scored by an administrator following a comprehensive interview with each respondent as well as an extensive file review. Studies such as these have offered significant contributions and insight into the current conceptualization of adolescent psychopathy. In an effort to extend the empirical base regarding test and item functioning of psychopathy assessment measures, the present study employed similar methods to investigate the functioning of a self-report measure of psychopathy-linked traits in adolescents.

Specifically, IRT analyses were utilized to investigate item functioning of the Inventory of Callous-Unemotional Traits (ICU; Frick, 2003). The ICU's unique emphasis on the affective features of psychopathy (i.e., callousness, uncaring, and unemotionality) facilitated a closer examination of these traits than previous measures have allowed. Callousness, uncaring, and unemotionality are each operationalized through multiple items as opposed to previous measures, such as the PCL:YV and APSD, which rely on fewer items as indicators of CU traits. As discussed above, previous research has provided evidence suggesting items assessing callousness or a lack of empathy tend to perform superior to others, such as those measuring behavioral features, in terms of discriminating power and amount of information provided. Therefore, it appeared worthwhile to investigate these affective traits more specifically through IRT analyses to explore possible differences in item functioning as well as the utility of the ICU in a sample of adolescent offenders and non-offenders.

The current study investigated which CU traits as assessed by the ICU were the most useful and meaningful regarding assessment of psychopathy-linked traits in adolescents. This study also explored which, if any, items were particularly useful for discriminating among adolescents at the highest or most severe levels of the underlying construct of psychopathy. Implications for assessment and the application of these findings in clinical and forensic settings were of particular interest to this study. Lastly, this study utilized IRT methods to explore the possibility of differential item functioning between adolescent offenders and non-offenders. Determining which items are more useful for identifying psychopathy-linked characteristics in adolescent offenders as compared to non-offenders offers insight into traits particularly salient for early identification of, and intervention for, adolescents at risk for developing persistent patterns of maladaptive or antisocial behavior.

Previous researchers (e.g., Edens et al., 2001; Moffitt, 2003) suggest that some of the characteristics commonly associated with psychopathy are part of normal adolescent development and not necessarily indicative of severe personality or behavioral deficits. Traits such as irresponsibility, impulsivity, and grandiosity are presumably common during adolescence and are likely to be transient characteristics expected to dissipate as adolescents transition into adulthood. Including items assessing traits common to most adolescents could potentially inflate psychopathy scores resulting in over-pathologizing of some individuals. Based on these concerns, the current study investigated which CU traits as assessed by the ICU are commonly endorsed in both groups of adolescents, suggesting the possibility that some traits may be common during adolescence and not necessarily indicative of maladaptive functioning.

It was predicted that items of the ICU tapping into callousness, followed by items assessing uncaring, would provide the highest levels of discrimination as well as item information regarding the underlying construct of psychopathy (Hypothesis 1) based on the associations of the ICU Callousness and Uncaring scales with problem behaviors in previous studies (Essau et al., 2006; Kimonis, Frick, Skeem et al., 2008). It was further predicted that items assessing callousness would be particularly useful for the assessment of psychopathy-linked traits at the highest levels of the trait (Hypothesis 2). In other words, the probability of endorsing those items would be greatest at the highest levels of theta, whereas there would be a low probability that callousness items would be endorsed at lower levels of the construct. Lastly, it was expected that differential item functioning would exist between a detained sample of adolescents (i.e., offenders) and a community adolescent sample operationalized as non-offenders (Hypothesis 3). Items exhibiting differential item functioning for adolescent offenders may warrant special attention as risk factors for maladaptive behavioral patterns as well as significant emotional deficits.

CHAPTER III

METHODOLOGY

Participants

Participant data was obtained from three previously collected data sets. The sample operationalized as non-offenders consisted of 166 male and female adolescents ranging in age from 14 to 18 years. Data were collected from these adolescents as part of a larger research study (Marsee, 2008). All of these individuals were enrolled in 9th through 12th grade at two public high schools in south Mississippi. Although the specific presence or absence of a history of illegal offenses for these adolescents is unknown, participants' enrollment and attendance in school are considered indicators of a lack of, or at most a minimal amount of, legal system involvement. Data from 15 participants were excluded from the current study due to one or more missing item responses on the ICU. The final non-offender sample consisted of 151 adolescents (39.7% male) ranging in age from 14 to 18 years ($M = 14.96$, $SD = 1.07$).

The offender group consisted of two independent samples, combined for the purposes of the present study, of male and female adolescents detained in juvenile justice settings. The detained female sample (Marsee & Frick, 2007) consisted of 88 pre-adjudicated adolescent girls ranging in age from 12 to 18 ($M = 14.98$, $SD = 1.30$). At the time of data collection, the females were awaiting trial in three separate short-term detention centers located in southern Louisiana. According to information obtained during file reviews, 35% of this sample has a prior history of at least one violent offense. Histories of nonviolent offenses committed by this sample include public order offenses,

status offenses, and theft (59%, 22%, and 21%, respectively) with some individuals committing multiple types of offenses (Marsee & Frick, 2007).

Dr. Eva Kimonis provided data from a sample of detained males (see Kimonis, Frick, Munoz, & Aucoin, 2008). This sample consisted of 88 male adolescents ranging in age from 13 to 18 years ($M = 15.57$; $SD = 1.28$) who were placed in a juvenile detention center in the southeastern United States. According to information obtained through a file review, approximately half (51%) of this sample has a positive history for at least one violent offense. At the time of data collection, offenses included “violent (30.7%), property (40.9%), status (9.1%), drug (11.4%), and other types” (Kimonis, Frick, Munoz et al., 2008; p. 573). Among the 176 male and female offenders, 18 were excluded from the current study due to one or more missing item responses on the ICU. The final offender sample consisted of 158 adolescents (62% male) ranging in age from 12 to 18 years ($M = 15.29$, $SD = 1.3$).

Therefore, the total sample for this study was comprised of $N = 309$ (158 offenders, 151 non-offenders) adolescents (51.3% male) ranging in age from 12 to 18 years ($M = 15.14$, $SD = 1.21$).

Materials

Demographic Information

Demographic information was collected as part of the assessment battery for each of the three groups. For the present study, demographic variables of interest were limited to age and sex, mainly for descriptive purposes.

Inventory of Callous-Unemotional Traits (ICU; Frick, 2003)

The items comprising the ICU were derived from the Antisocial Process Screening Device (APSD; Frick & Hare, 2002), particularly the APSD CU scale. Based on factor analyses of the APSD revealing four items that consistently loaded onto the APSD CU scale, 24 items were developed and included on the ICU. Each item is rated using a four-point Likert scale from 0 indicating *not at all true* to 3 indicating *definitely true*. The ICU offers a multidimensional conceptualization of CU traits by including several items assessing each of the three aspects of callousness, uncaring, and unemotionality. The ICU has demonstrated an acceptable overall internal consistency ($\alpha = .77$; Essau et al., 2006) in previous research. Furthermore, acceptable internal consistencies have been found for the Callousness and Uncaring subscales (.70 and .73, respectively), although the Unemotional subscale demonstrated only modest internal consistency ($\alpha = .64$; Essau et al., 2006). In a separate study examining the reliability of the ICU subscales in an at-risk adolescent sample, modest coefficient alpha levels of .65 were found for both the Callousness and Unemotional subscales, whereas the Uncaring subscale demonstrated an acceptable internal consistency of .79 (Ansel, 2009). In the current study, acceptable internal consistency estimates were found for both the offender and non-offender samples ($\alpha = .74$ and .83, respectively).

Procedure

Non-Offending Adolescents

Data collection for the non-offending group took place between December 2006 and March 2007. Following approval from the Institutional Review Board (IRB) at The University of Southern Mississippi, parental consent forms were collected for potential

participants. Only those students with parental consent were allowed to participate in the research study. Students were provided with a description of data collection procedures and for students who agreed to participate, signed assent forms were collected. The participants then completed the assessment battery, lasting approximately 60 to 90 minutes, which included several measures as part of a larger study (Marsee, 2008). Only the ICU and demographic information were of interest to the current study.

Female Offenders

After obtaining IRB approval and parental consent, each potential participant was informed of data collection procedures and given the opportunity to agree or refuse to participate. An assent form was signed by those females who agreed to participate in the study. Questionnaires were then administered in small groups (Marsee & Frick, 2007). Though several measures were collected during this time, only the ICU and demographic information were of interest to the current study.

Male Offenders

After obtaining parental consent to participation, those adolescents who had received permission from their parents to participate were provided with an explanation of the data collection procedures. Adolescents who agreed to participate signed an assent form prior to data collection. Data collection included completion of several questionnaires as well as a computer task (Kimonis, Frick, Munoz et al., 2008). Only the ICU and demographic data were of interest to the current study.

CHAPTER IV

ANALYSIS OF DATA

Assumptions of IRT

As previously mentioned, the assumptions of unidimensionality and local independence of a test should be met to appropriately conduct IRT analyses.

Confirmatory factor analyses were utilized to obtain estimates of unidimensionality for each group (i.e., offender, non-offender). If an acceptable degree of unidimensionality is established, it may be concluded that the assumption of local independence has been met (Hambleton et al., 1991). As described previously, local independence asserts that each item is only correlated with all other items of a test because of the underlying trait.

Therefore, if a test is unidimensional, the individual item responses will be inherently dependent on the one overall dimension or trait. If the influence of that trait is controlled for, then the item responses would be statistically random indicating local independence (Hambleton et al., 1991).

IRT Model

Based on the nature of the ICU, which relies on an ordered-categorical scale for scoring (i.e., item responses ranging from 0 to 3), and consistent with previous research applying IRT to psychopathy instruments (Schrum & Salekin, 2006; Vincent, 2002), Samejima's GRM (Samejima, 1969) was applied using Multilog 7 (Thissen, Chen, & Bock, 2003). This model is appropriate when the test instrument has ordered responses for which the item score is expected to increase as the level of θ increases. Theoretically, as the level of θ increases, the probability of a 0 response will decrease, whereas the probability of a 3 response will increase (Cooke & Michie, 1997). Since the ICU has four

response options, the probability of a 1 or 2 response should increase and then decrease as the level of θ increases from small to moderate, then from moderate to high trait levels, with the probability of a 1 response peaking at a lower trait level than a 2 response.

To examine the first hypothesis which predicted that items assessing callousness would produce the highest discrimination levels as well as provide the most information at the item level, item parameters for all of the ICU items were examined. The a parameter provides an index of item discrimination. As the a parameter increases, the slope becomes steeper indicating higher discrimination. Graphing the a and b parameters for each item provided separate Item Characteristic Curves (ICCs) for each of the items. The ICCs for each item were investigated as an index of item difficulty as well as item information. Those items producing the steepest slopes were considered the most useful discriminators.

The ICCs were also used to determine whether items assessing callousness were especially useful at the highest levels of psychopathy (Hypothesis 2). When graphed, ICCs that are shifted toward the highest levels of θ with the steepest slopes are indicative of items that differentiate only at the highest trait levels (Vincent, 2002). It was expected that the ICCs of callousness items would behave in such a way to suggest that only individuals possessing the highest levels of psychopathy-linked characteristics would endorse high responses on these items. For this hypothesis to be supported, ICCs of callousness items need to be primarily located along the highest levels of θ , whereas the remaining items should have yielded ICCs within the lower and moderate levels of the underlying trait.

In order to test whether any of the ICU items were demonstrating DIF between adolescent offenders versus non-offenders (Hypothesis 3), it was necessary to standardize the b parameters from each group to obtain a common metric for which to compare the groups (Hambleton et al., 1991). Comparisons of the item parameters defining the ICCs were used to detect DIF between groups. DIF would be evidenced if the ICC parameters of particular items were not equivalent across both groups. Nonuniform DIF was of particular interest to the present study. Nonuniform DIF refers to significant group differences specifically in the item discrimination parameters. As described by Vincent (2002) regarding age-related bias, the presence of nonuniform DIF in the current study would indicate that certain CU traits as assessed by the ICU were more or less relevant to the underlying trait depending on offender status.

Descriptive Statistics

Descriptive statistics for the offender and non-offender samples are provided in Table 1. The total ICU scores for the entire sample appear to be normally distributed ($skewness = .35$, $SE = .14$) with a *mean* ICU total score of 23.76 ($SD = 9.12$). The ICU scores demonstrated adequate internal consistency in both the offender and non-offender samples ($\alpha = .74$ and $.83$, respectively). Therefore, use and interpretation of ICU responses for the purposes of the current study appears appropriate.

Factor Analyses

Amos 19 (Arbuckle, 2010) was used to conduct a CFA of the ICU. Based on previously published factor loadings (Essau et al., 2006; Kimonis, Frick, Skeem et al., 2008), a three-factor bifactor model was tested to assess dimensionality.

Table 1

Demographic and Descriptive Statistics

Participants	N	Mean (Std Dev)	Age Range	Gender (% male)	
Total	309	15.14 (1.21)	12-18 years	51.3%	
Offender	158	15.29 (1.3)	12-18 years	62%	
Non-offender	151	14.96 (1.07)	14-18 years	39.7%	
ICU Total and Subscale Scores	<i>M</i>	<i>SD</i>	Minimum (possible = 0)	Maximum (possible = 72)	Skewness (Std Error)
Overall Total	23.76	9.12	1	48	.35 (.14)
Offender Total	23.63	8.64	1	48	.10 (.19)
Callousness	8.54	4.71			
Uncaring	9.22	4.94			
Unemotional	7.42	3.09			
Non-offender Total	23.89	9.57	6	47	.54 (.20)
Callousness	6.99	4.28			
Uncaring	9.21	4.81			
Unemotional	7.69	3.19			
Subscale Correlations	<i>r</i> (<i>p</i> < .01)				
Callousness – Uncaring	.36				
Callousness – Unemotional	.19				
Uncaring – Unemotional	.26				

With regard to the ICU, the three-factor bifactor model proposes that a general CU factor accounts for responses to each of the ICU items. In addition, three independent subfactors exist, corresponding to the three ICU subscales (i.e., Callousness, Uncaring, and Unemotional), which account for unique variance in their identified set of items. The unique variance explained by the subfactors is thought to be over and above that accounted for by the general CU factor. For this model, the subfactors are assumed to be uncorrelated (Kimonis, Frick, Skeem et al., 2008). Several fit indices were utilized when

interpreting results of model fit due to the absence of an agreed upon “best indicator” in the research literature (Schumacker & Lomax, 2004). Indices utilized include chi-square (X^2), comparative fit index (CFI), and the root mean squared error of approximation (RMSEA).

First, the fit of the three-factor bifactor model was tested using the overall sample data ($N = 309$; see Table 2 for the CFA results). The results indicated an inadequate model fit to the data according to the X^2 and CFI criteria ($df = 228$, $X^2 = 488.41$, $p < .001$, CFI = .83). However, the obtained RMSEA value of .06 is considered to be indicative of adequate model fit (Browne & Cudeck, 1993).

Table 2

Confirmatory Factor Analyses (CFA) Results

Model	<i>df</i>	Chi-square (X^2)	CFI	RMSEA	X^2/df
1. 3-factor bifactor with all ICU items	228	488.41*	.83	.06	2.14
2. 3-factor bifactor with items 2 and 10 excluded	187	429.7*	.83	.07	2.30
3. Multigroup Invariance Unconstrained	456	764.15*	.81	.05	1.68
4. Multigroup Invariance Structural covariances	504	860.93*	.78	.05	1.71

Note. * $p < .001$; *df* = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

An analysis of the item-total correlations revealed that ICU items 2 and 10 (i.e., “What I think is right and wrong is different from what other people think” and “I do not let my feelings control me”) were not significantly correlated with total ICU scores ($r_s = .09$ and $.05$, respectively). These results are consistent with previous findings regarding

the ICU items (Kimonis, Frick, Skeem et al., 2008). These two items were deleted from the three-factor bifactor model, and a second CFA was conducted to determine the degree of model fit without these seemingly unrelated items. Deleting these two items did not appear to improve model fit. With items 2 and 10 excluded from the model, an inadequate model fit was maintained by the X^2 and CFI results ($df = 187$, $X^2 = 429.7$, $p < .001$, CFI = .83). Consistent with the first model, the fit is considered adequate according to the obtained RMSEA fit index of .07.

The discrepancies between fit indices found in the current study are consistent with the pattern of results obtained in a previous CFA of the ICU items (Kimonis, Frick, Skeem et al., 2008). However, the offender sample data used in the current study was taken from the participant data utilized by Kimonis, Frick, Skeem, et al. (2008). Therefore, 51% ($N = 158$) of the current participant data overlap with the data used in the previous factor analysis.

To determine whether this model yielded equivalent fit for the offender and non-offender samples, a test of cross-group invariance was utilized. For this analysis, both groups were fitted simultaneously to an unconstrained model, $X^2 = 764.15$ ($df = 456$, $p < .001$). The obtained result was then compared to a restrained model, $X^2 = 860.93$ ($df = 504$, $p < .001$). A significant difference was found between groups ($X^2 = 96.78$, $df = 48$, $p < .001$). A significant change in model fit is indicative of unequal model fit across groups. Results of these analyses are shown in Table 2.

Based on the factor analyses results as well as the nature of the ICU item response data, an exploratory factor analysis (EFA) was conducted using Mplus 5.1 (Muthen & Muthen, 2006). Mplus allows for analyses which take into account the categorical nature

of the ICU data as opposed to Amos which assumes that the observed variables (i.e., item responses) are continuous. An EFA was conducted using the total combined sample to determine if differences in factor structure emerged when the variables were redefined, more appropriately, as categorical. Items with a .30 or higher factor loading were retained. Results are provided in Table 3.

Table 3

Exploratory Factor Analysis (EFA) Results

Geomin Factor Loadings			
ICU Items	Factor 1	Factor 2	Factor 3
1	.021	.383	.556
2	.079	.111	-.318
3	.618	.102	-.031
4	.510	.165	-.281
5	.630	-.081	.089
6	-.028	.566	.325
7	.379	.482	.018
8	.539	.156	.040
9	.482	.384	-.133
10	-.137	.101	.079
11	.489	.583	.011
12	.403	.418	-.103
13	.260	.005	.360
14	.075	-.023	.627
15	.694	-.056	.047
16	.645	-.024	.274
17	.682	-.077	.004
18	.371	.444	-.124
19	.158	.097	.639
20	.297	.574	.015
21	.337	.464	-.027
22	-.021	.663	.338
23	.649	-.024	.164
24	.488	-.041	.200

Note. **Bold** numbers indicate items retained for each factor. *Italics* numbers indicate items with factor loadings $\geq .30$.

Consistent with the factor structure previously presented, three factors were identified. However, the items loaded somewhat differently among the three factors. The first factor consisted of ten items and was most consistent with the original Uncaring subscale. Eight out of the ten items on Factor 1 are reverse-scored. The second factor, most similar to the Callousness subscale, consisted of eight items none of which are reverse-scored. Finally, Factor 3 consisted of four items, all of which are reverse-scored, and closely resembled the original Unemotional subscale. Consistent with previous results, Items 2 and 10 performed poorly and did not load on any of the factors with a .30 or higher loading. The EFA results suggest that item-format (i.e., reverse-scoring) may be the primary influence in the obtained factor structure rather than shared item content or similarity of traits assessed.

IRT Analyses

Despite the weak factorial invariance across groups demonstrated in this sample, all ICU items were retained in the IRT analyses based on the overall rationale for conducting this study. Since standard administration of this measure includes all items, it seemed inappropriate to exclude any items from the primary statistical analyses. Furthermore, there did not appear to be an obvious “worst item” that, if deleted, would significantly improve model fit. Therefore, several items would have needed to be deleted which would have detracted from the overall purpose of the current study. In short, the results of the IRT analyses should be interpreted with caution in light of the findings from the factor analyses, but consideration of the full scale is considered appropriate based on the purpose and exploratory nature of the current study.

Item Discrimination and Item Difficulty Parameters

Samejima's GRM (Samejima, 1969) was used to investigate most aspects of item functioning with Multilog 7 (Thissen, Chen, & Bock, 2003). The GRM is appropriate for polytomous test items yielding ordered-categorical response data. First, separate analyses were conducted for each group (i.e., offender and non-offender) to determine item functioning for each sample independently. Item parameter estimates for the offender and non-offender samples are presented in Table 4.

Table 4

Item Parameter Estimates for Independent Samples

	Offender (n=158)				Non-offender (n=151)			
	<i>a</i>	<i>b₁</i>	<i>b₂</i>	<i>b₃</i>	<i>a</i>	<i>b₁</i>	<i>b₂</i>	<i>b₃</i>
Callousness								
Item 2	.16	-8.33	3.16	9.28	.25	-6.94	2.61	7.60
Item 4	.62	1.85	3.83	5.77	1.74	1.87	2.70	2.87
Item 7	.53	1.22	3.74	5.30	1.65	.75	2.27	2.60
Item 8	1.13	-1.60	.17	2.27	1.27	-.88	.65	2.32
Item 9	.77	1.02	2.74	3.50	1.60	.21	1.28	2.05
Item 10	.15	-6.90	3.75	10.31	.18	-6.97	3.68	10.61
Item 11	.99	1.13	2.55	3.28	1.87	1.08	1.99	2.34
Item 12	.51	.77	3.39	5.91	1.28	1.00	2.23	2.68
Item 18	.77	.11	1.96	3.77	.79	.31	2.90	4.07
Item 20	.46	.89	4.04	6.60	1.26	.58	2.23	3.30
Item 21	.64	.35	2.48	4.28	.95	1.08	2.57	3.70
Uncaring								
Item 3	1.20	-.13	1.05	2.54	1.74	.14	1.19	2.39
Item 5	1.62	-.52	.46	1.81	1.33	-1.00	.42	2.09
Item 13	1.03	-1.48	-.10	1.42	.59	-4.27	-1.79	1.81
Item 15	1.65	-.26	.74	2.12	2.18	-.32	.41	2.07
Item 16	1.50	-.77	.20	2.15	1.84	-.58	.34	1.79
Item 17	1.64	-.78	.35	2.21	1.42	-.81	.59	1.83
Item 23	1.88	-.46	.52	1.70	1.86	-.70	.35	1.92
Item 24	1.34	-1.14	-.03	1.89	1.04	-1.34	.38	3.07
Unemotional								
Item 1	.56	-2.94	-1.08	2.45	.45	-3.74	-1.89	3.86
Item 6	.18	-5.00	3.31	8.76	.54	-1.99	2.15	4.17
Item 14	.44	-3.86	-.59	2.29	.51	-3.90	-2.03	1.74

Table 4 (continued).

	Offender (n=158)				Non-offender (n=151)			
	<i>a</i>	<i>b₁</i>	<i>b₂</i>	<i>b₃</i>	<i>a</i>	<i>b₁</i>	<i>b₂</i>	<i>b₃</i>
Item 19	.69	-2.37	-.36	1.74	.76	-2.60	-1.45	.82
Item 22	.16	-5.51	3.55	8.78	.80	-1.67	1.16	2.49

Note. Item discrimination parameters = *a*; Item difficulty, or threshold, parameters = *b₁*, *b₂*, and *b₃*.

The ICCs and item parameters for each item of the ICU were investigated to test whether items assessing callousness, followed by items assessing uncaring, produced the highest discrimination levels as well as provided the most information at the item level

(Hypothesis 1). The ICU items are provided in Appendix A. In the non-offender sample, Item 15 (i.e., “I always try my best” reverse scored) was the most discriminating item ($a = 2.18$) followed by Items 11 ($a = 1.87$), 23 ($a = 1.86$), and 16 ($a = 1.84$) [i.e., “I do not care about doing things well,” “I work hard on everything I do” reverse scored, and “I apologize (say I’m sorry) to persons I hurt” reverse scored]. Three of these items (15, 23, and 16) belong to the Uncaring subscale, whereas Item 11 assesses callousness.

Regarding discriminating items in the offender sample, Item 23 ($a = 1.88$) yielded the highest discrimination parameter, followed by Items 15 ($a = 1.65$), 17 ($a = 1.64$; i.e., “I try not to hurt others’ feelings” reverse scored), and 5 ($a = 1.62$; i.e., “I feel bad or guilty when I do something wrong” reverse scored). All of these items belong to the Uncaring subscale of the ICU. Item 10, an item on the Callousness subscale, yielded the lowest a parameter for both the offender and non-offender samples ($as = .15$ and $.18$, respectively) and was therefore the least discriminating item across the two samples. This finding is not surprising based on item-total correlations discussed previously. Items 15 and 23 from the Uncaring subscale were among the most discriminating items across both

samples, indicating that based on examination of ICCs, Hypothesis 1 was not fully supported.

It is necessary also to consider the item difficulty parameters (b s) when determining overall item information. For each of the ICU items, three difficulty parameters were estimated (i.e., b_1 , b_2 , and b_3). The difficulty, or threshold, parameters represent the thresholds between the four response items. Difficulty parameters can be conceptualized as a series of dichotomies. For example, b_1 represents the level of θ (CU traits) necessary to have a .5 probability of responding above the lowest category (i.e., 0 versus 1, 2, and 3) whereas b_2 represents the level of θ necessary to endorse the two highest response categories versus the two lowest categories (i.e., 0, 1 versus 2, 3) with a .5 probability. Refer to Table 4 for difficulty parameter estimates for both samples. For the purposes of the current study, b_3 estimates were examined for items demonstrating the highest discrimination parameters to determine overall item information because b_3 represents the level of the underlying CU trait necessary to endorse the highest response category for a given item with a .5 probability. In the non-offender sample, Item 11 from the Callousness subscale yielded the highest difficulty parameter ($b_3 = 2.34$) among those items with the highest discrimination parameters. In the offender sample, Item 17 (Uncaring subscale) appeared to be the most difficult ($b_3 = 2.21$) among those items producing the highest discrimination parameters. These results indicate that Items 11 and 17 are particularly useful for identifying individuals with higher levels of CU traits from the respective samples. Additionally, the elevated discrimination parameters suggest that individuals with lower levels of CU traits would be unlikely to endorse these items. In this regard, the first hypothesis appears to be partially supported for the non-offender

sample based on the discrimination and difficulty parameters produced by Item 11.

However, Uncaring items produced the greatest item discrimination and difficulty parameters for the offender sample. As expected, items from the Unemotional subscale were the least discriminating (lowest a parameters) for both samples (see Table 4).

Item Functioning at the Highest Levels of Theta

It was further hypothesized that items assessing callousness would be particularly useful for the assessment of CU traits at the highest levels of the underlying construct (Hypothesis 2). Based on the location of ICCs of Callousness items compared to ICCs of Uncaring items in both samples, this hypothesis appears to have been partially supported. ICCs from each sample are presented in Figures 1 and 2. Overall, the ICCs of Callousness items in both samples were primarily shifted toward the highest levels of θ . Graphically speaking, these ICCs are primarily located on the right side of the figures. There appeared to be greater variability among the Uncaring items based on ICCs spanning a wider range of θ . In other words, as shown in Figures 1 and 2, the Uncaring items yielded ICCs with locations ranging from moderate to high levels of θ , whereas the Callousness ICCs were primarily located at the highest levels of θ . Despite yielding ICCs along the highest levels of θ , the Callousness items were generally less discriminating in both samples. Therefore, their locations along θ may be less useful because these items demonstrated inferior discriminating abilities compared to most of the Uncaring items. Further interpretation of these results is provided in the Discussion section.

Estimates of overall test information for each sample were also examined. Though total test information was not a primary focus of this study, examinations of these estimates enhance the interpretation of the results.

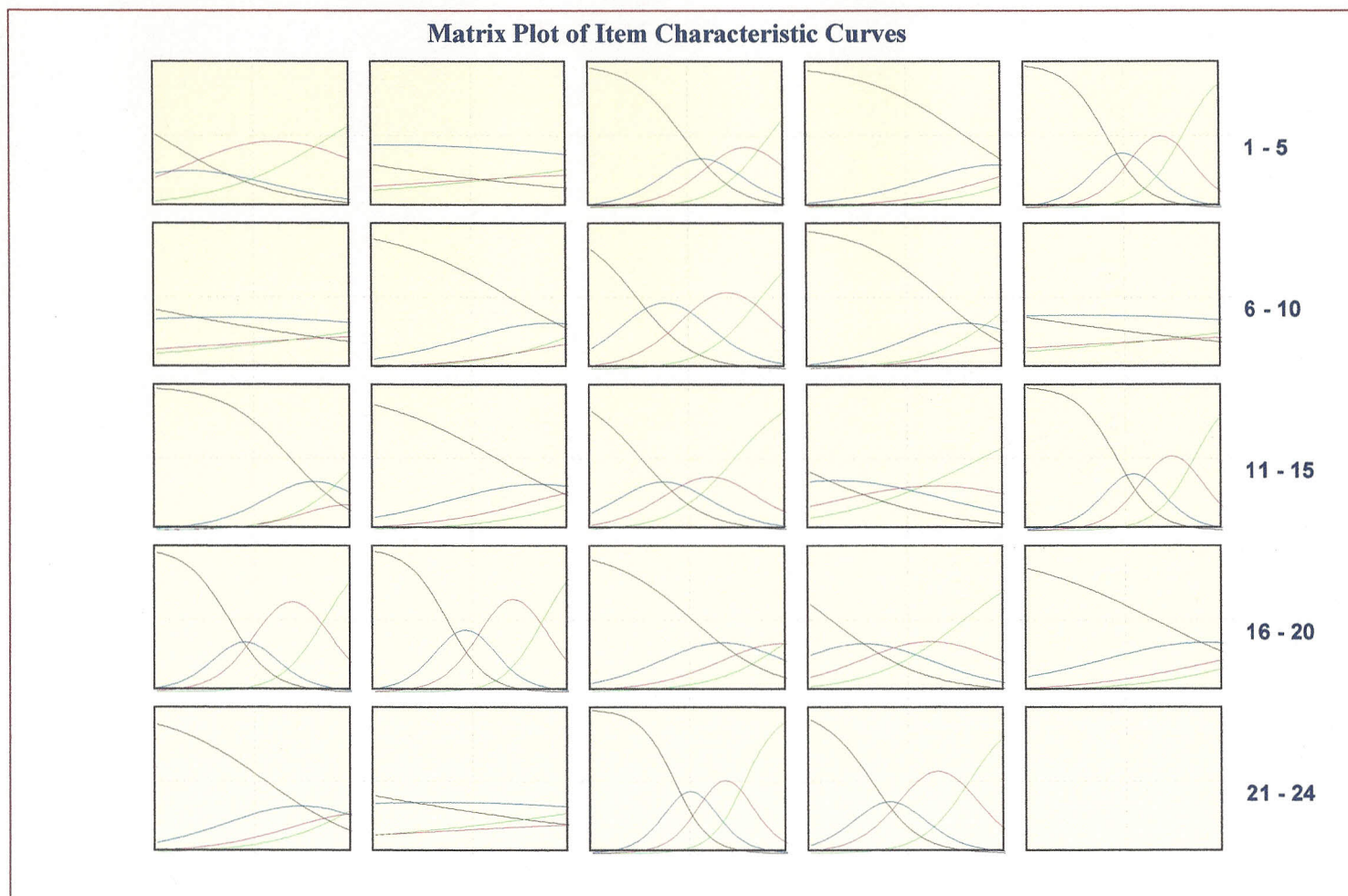


Figure 1. Item Characteristic Curves for ICU Items in Offender Sample. Colored lines represent the Category Response Curves for each item (Black = 1, Blue = 2, Pink = 3, Green = 4). The x-axis is the underlying CU trait (θ) ranging from -3 to 3 with the vertical dotted line indicating a trait level of 0. The y-axis is the probability of a response with the horizontal dotted line representing a .5 probability of a response at a given level of θ .

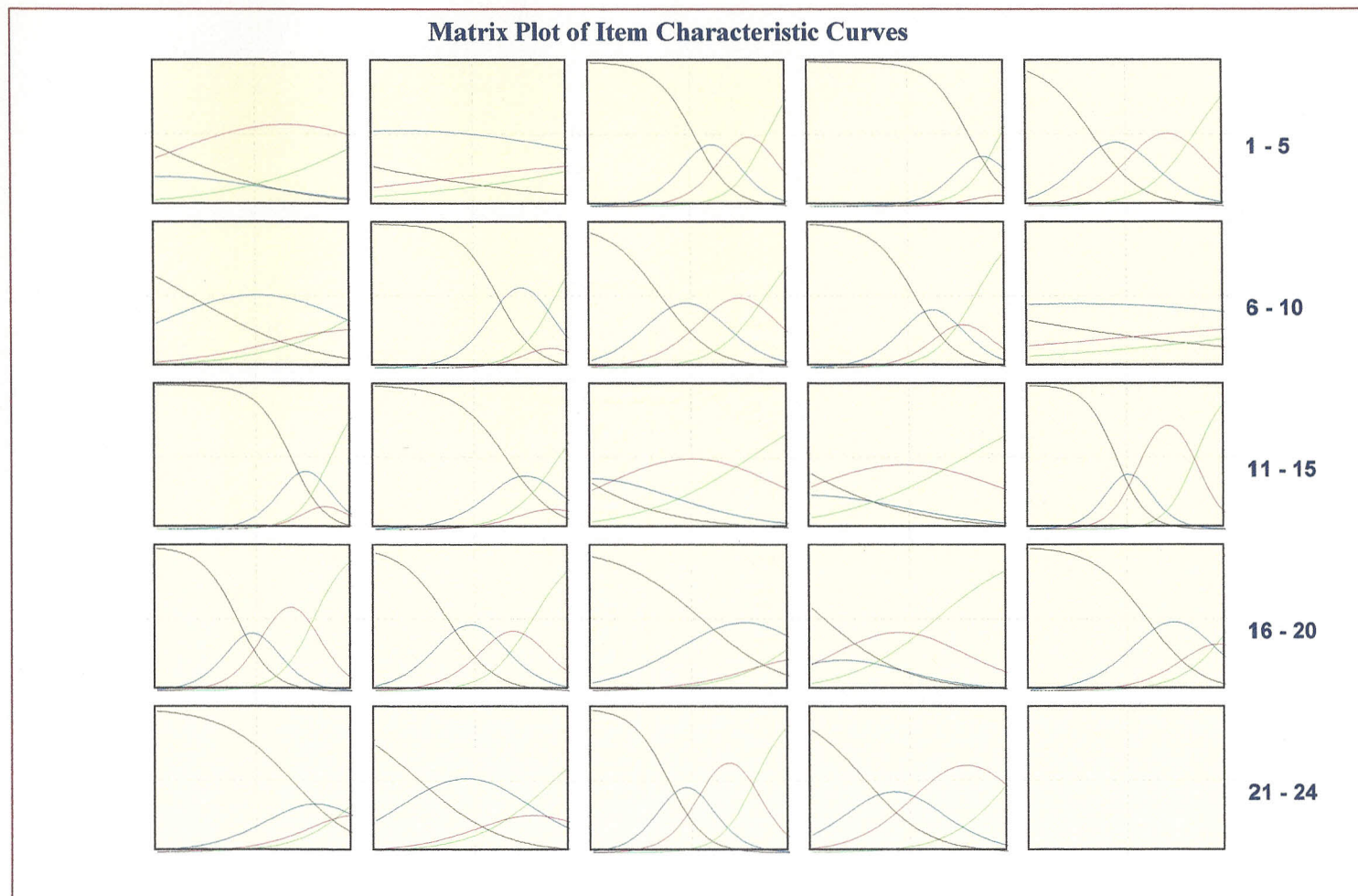
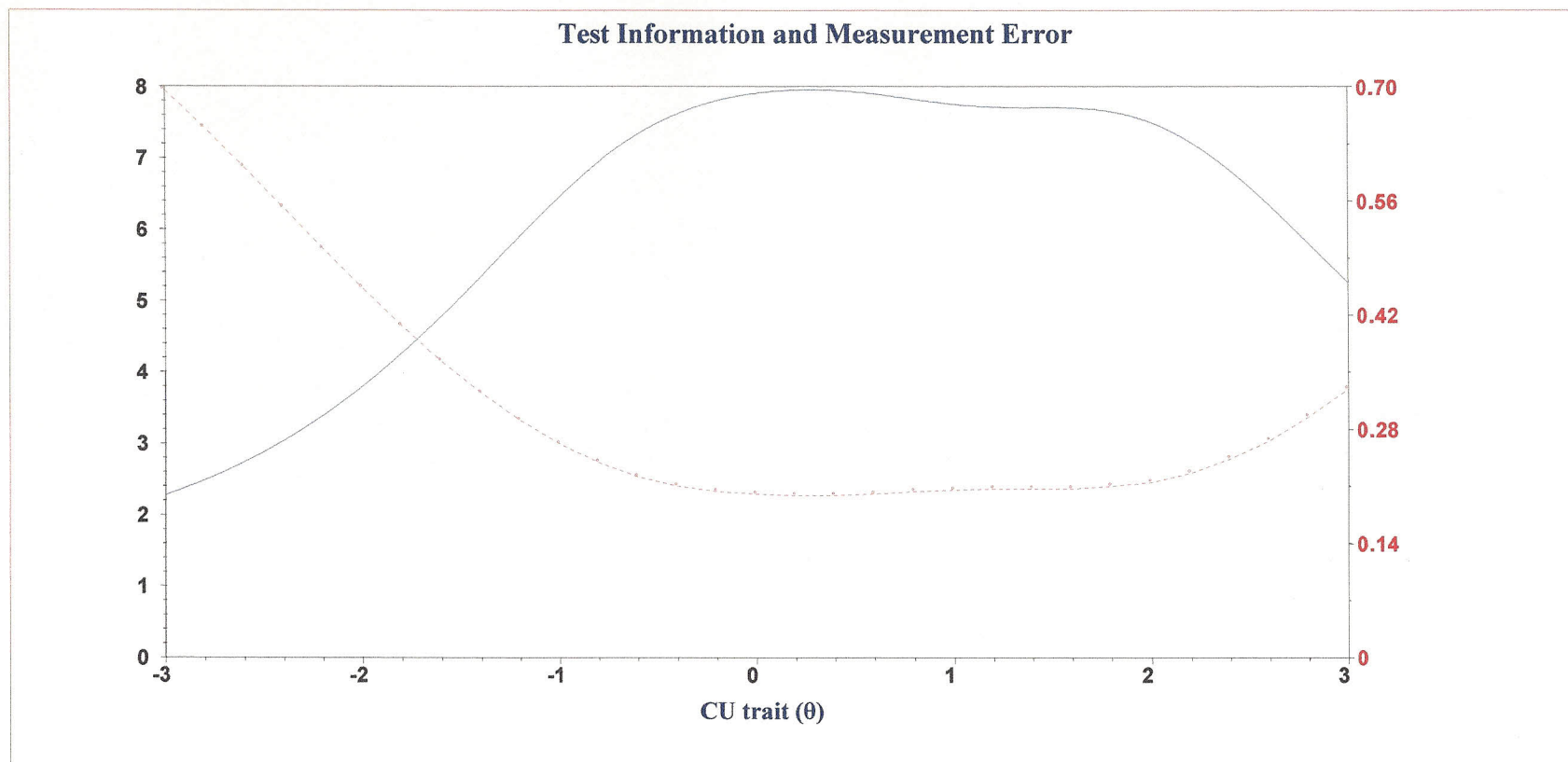


Figure 2. Item Characteristic Curves for ICU Items in Non-Offender Sample. Colored lines represent the Category Response Curves for each item (Black = 1, Blue = 2, Pink = 3, Green = 4). The x-axis is the underlying CU trait (θ) ranging from -3 to 3 with the vertical dotted line indicating a trait level of 0. The y-axis is the probability of a response with the horizontal dotted line representing a .5 probability of a response at a given level of θ .

Refer to Figures 3 and 4 for illustrations of total test information by sample. In the offender sample, the ICU provided the most information regarding the underlying construct for moderate levels of CU traits (θ s ranging from .2 to .4). In other words, the ICU was the most reliable in estimating CU traits in the offender sample for moderate trait levels. In the non-offender sample, the ICU provided the most information at higher trait levels (θ s ranging from 1.8 to 2.0).

Evaluation of Differential Item Functioning (DIF)

This study also investigated whether ICU items functioned differently in a non-offender adolescent sample compared to a sample of adolescent offenders. It was hypothesized that DIF would exist between these two adolescent samples (Hypothesis 3), although no predictions were made regarding particular items based on the exploratory nature of this study. This hypothesis was supported. Of particular interest to the current study was non-uniform DIF which refers to significant differences among the item discrimination parameters based on group membership. The non-offender (community) sample was designated as the reference group, whereas the offender sample was considered the focal group for this study. To test for DIF, the two groups were first forced onto a common metric which allows for a standardized comparison of parameters. An unconstrained model was then estimated which imposes no equality restraints on the item parameters across groups. Next, the results of the fit assessment from the unconstrained model were compared to a model with item discrimination parameters (a parameters) constrained to be equal. In this constrained model, the b parameters for each group were estimated with the discrimination parameters set to be equal across groups (see Table 5). G^2 was used as a goodness-of-fit index to evaluate the difference in fit assessment indices (i.e., -2 times the log-likelihood) produced by the two models, with degrees of freedom



Test information curve: solid line

Standard error curve: dotted line

The total test information for a specific scale score is read from the left vertical axis.

The standard error for a specific scale score is read from the right vertical axis.

Figure 3. Total Test Information for Offender Sample.

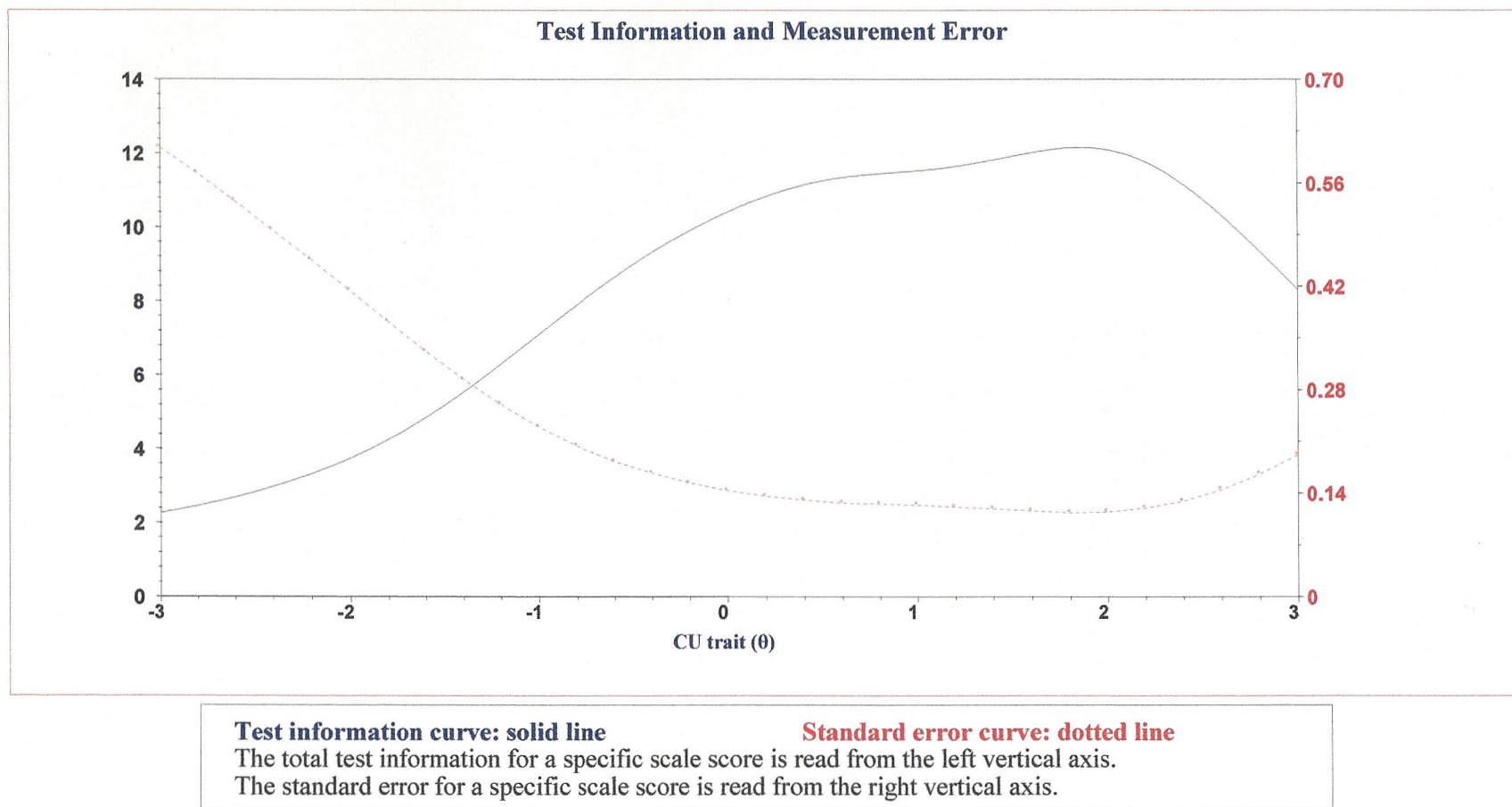


Figure 4. Total Test Information for Non-Offender Sample.

equal to the number of additional parameter constraints in the constrained model versus the unconstrained model. The difference between the unconstrained model and the model with all discrimination (a) parameters constrained to be equal was significant, $G^2(24) = 50.8, p < .01$. A significant G^2 provides support for the hypothesis that DIF exists, specifically non-uniform DIF, suggesting that ICU items functioned differently in this sample depending on group classification.

Table 5

Item Parameters for Offender Sample Alone and Combined Offender and Non-Offender Model With Common Metric

	Prior to Common Metric				Common Metric						
	Offenders				Non-Offenders				Offenders		
	a	b_1	b_2	b_3	a	b_1	b_2	b_3	b_1	b_2	b_3
Callousness											
Item 2	.16	-8.33	3.16	9.28	.29	-6.59	1.54	5.74	-4.66	1.82	5.15
Item 4	.62	1.85	3.83	5.77	1.16	1.73	2.92	3.18	1.06	2.15	3.21
Item 7	.53	1.22	3.74	5.30	1.14	.27	2.27	2.73	.63	1.87	2.62
Item 8	1.13	-1.60	.17	2.27	1.32	-1.57	-.05	1.60	-1.40	.12	1.92
Item 9	.77	1.02	2.74	3.50	1.30	-.45	.78	1.69	.65	1.73	2.20
Item 10	.15	-6.90	3.75	10.31	.24	-5.88	2.12	7.20	-4.11	2.35	6.26
Item 11	.99	1.13	2.55	3.28	1.60	.51	1.52	1.91	.77	1.72	2.20
Item 12	.51	.77	3.39	5.91	.99	.52	2.03	2.59	.42	1.82	3.15
Item 18	.77	.11	1.96	3.77	.93	-.42	1.87	2.87	.07	1.62	3.14
Item 20	.46	.89	4.04	6.60	.93	.04	2.16	3.57	.47	2.09	3.38
Item 21	.64	.35	2.48	4.28	.95	.41	1.91	3.04	.24	1.73	2.96
Uncaring											
Item 3	1.20	-.13	1.05	2.54	1.44	-.53	.66	2.04	-.12	.87	2.11
Item 5	1.62	-.52	.46	1.81	1.49	-1.65	-.30	1.29	-.54	.44	1.79
Item 13	1.03	-1.48	-.10	1.42	.82	-3.91	-2.06	.66	-1.76	-.15	1.64
Item 15	1.65	-.26	.74	2.12	1.77	-1.06	-.22	1.66	-.24	.66	1.89
Item 16	1.50	-.77	.20	2.15	1.71	-1.31	-.32	1.22	-.68	.16	1.84
Item 17	1.64	-.78	.35	2.21	1.59	-1.47	-.13	1.06	-.78	.31	2.13
Item 23	1.88	-.46	.52	1.70	1.73	-1.43	-.32	1.34	-.48	.49	1.66
Item 24	1.34	-1.14	-.03	1.89	1.23	-1.91	-.37	2.04	-1.18	-.04	1.93
Unemotional											
Item 1	.56	-2.94	-1.08	2.45	.60	-3.58	-2.18	2.24	-2.75	-1.03	2.25
Item 6	.18	-5.00	3.31	8.76	.50	-2.83	1.62	3.80	-1.87	1.17	3.16
Item 14	.44	-3.86	-.59	2.29	.53	-4.52	-2.68	1.01	-3.26	-.53	1.89
Item 19	.69	-2.37	-.36	1.74	.75	-3.34	-2.17	.14	-2.20	-.36	1.56
Item 22	.16	-5.51	3.55	8.78	.56	-2.94	.93	2.75	-1.66	1.06	2.59

Note. Item discrimination parameters = a ; Item difficulty, or threshold, parameters = b_1 , b_2 , and b_3 .

For exploratory purposes, another model was estimated with the difficulty (b) parameters constrained to be equal while imposing additional constraints of three designated anchor items. Items 2, 6, and 23 were chosen as anchor items based on the results of the item parameter estimation analyses conducted prior to this analysis. Specifically, these items appear to represent ICU items with little to no DIF, spanning a wide range of the trait. The fit assessment results from this model were compared to the original unconstrained model to determine whether uniform DIF existed among the difficulty parameters. Uniform DIF refers to significant group differences in item difficulty parameters. A significant difference would indicate that the point along θ at which specific items, or CU traits, become useful differs depending on group classification. The difference between the unconstrained model and the model with all difficulty (b) parameters constrained to be equal was significant, $G^2(9) = 20.5, p < .05$, indicating the presence of uniform DIF.

IRTLRDIF 2 (Thissen, 2001) was used to obtain G^2 estimates for individual items when item parameters were estimated for both groups simultaneously. This procedure allows for the investigation of which specific items are functioning differently between samples. When testing the null hypothesis that corresponding items between the two groups would exhibit no DIF (no equality constraints imposed), four items yielded significant G^2 estimates at the $p < .05$ level. The content of these four items includes caring about personal responsibility (i.e., being on time, staying out of trouble; Items 7 and 9, respectively), others' perceptions of the respondent (e.g., appearing uncaring to others; Item 12), and openness about feelings with others (Item 22). Four additional items [item content including attitude toward hurting others for personal gain (Item 4), ease of

admitting to being wrong (Item 13), and emotional expressiveness (Items 14 and 19)] yielded significant G^2 estimates at the $p < .01$ level. These eight items represent psychopathy-linked characteristics from all three scales of the ICU (i.e., Callousness, Uncaring, Unemotional).

The next step was to examine the G^2 estimates when equality constraints were imposed for these eight items to determine the nature of the observed differences (i.e., non-uniform or uniform DIF). When discrimination (a) parameters were constrained to be equal, items 4, 7, and 22 yielded significant G^2 estimates at the $p < .05$ level, indicating that these items did not discriminate across levels of θ in the same way for both groups. All three items yielded higher discrimination parameters in the non-offender sample compared to the offender sample. Items were then analyzed with both the difficulty (b) and discrimination (a) parameters constrained to be equal. The remaining five items (items 9, 12, 13, 14, and 19) as well as Item 4 demonstrated significant uniform DIF, suggesting that these items are useful at different levels of the underlying trait. Specifically, these six items demonstrated greater usefulness at higher levels of θ in the offender sample compared to non-offenders. Further investigation of how the items function differently and possible implications for the use of the ICU in different populations are provided in the Discussion section.

In summary, the results indicate that both non-uniform and uniform DIF exist between this sample of adolescent offenders and non-offenders. Certain items of the ICU appeared to function differently between the two groups in terms of discriminating abilities along levels of θ , as well as in the locations along θ at which certain items become useful.

CHAPTER V

DISCUSSION

The purpose of this study was to examine the usefulness of the Inventory of Callous-Unemotional Traits (ICU), particularly the individual ICU items, for assessing psychopathy-linked traits in offender and non-offender adolescent samples.

Investigations of the factor structure of the ICU in the current sample suggest variance between groups (i.e., offender and non-offender) on the ICU factors. Interestingly, the two groups did not differ significantly on mean total ICU scores or range and skewness of obtained scores. However, item analyses indicated that functional differences exist among individual ICU items which contributed to differential item parameter estimates between groups.

IRT Evaluations of Item and Test Functioning

Contrary to expectations, Callousness items were not the most discriminating items in this study. Several factors may have contributed to this finding. First, the confirmatory factor analysis results suggest weak fit of the current division of items into three subscales for this sample. Based on the exploratory factor analysis using categorical variables, the obtained subscales appear to primarily consist of ICU items with shared item formats (e.g., reverse-scored vs. regular-scored) rather than items assessing similar aspects of CU traits. These findings raise the question of whether the established subscales and their corresponding items represent distinct aspects of CU traits in the manner previously described (see Essau et al., 2006). It should be noted, however, that the most discriminating and difficult items in each sample consistently loaded on the given subscales for both the CFA and EFA. Second, callousness as measured by the ICU may be associated with problem behaviors but may not necessarily capture the core

aspects of CU traits. Specifically, previously found associations between the ICU Callousness subscale and problem behaviors may not be due to the presence of callous traits, per se, but may be more representative of broader interpersonal and behavioral deficits captured by the Callousness items. In other words, callous traits as operationalized by the ICU may be more closely tied to general disregard for others rather than specific affective psychopathy-linked traits. Additionally, the poor discriminating ability of these items supports the separation of psychopathy-linked traits into behavioral and affective features. The current results suggest that, although callousness is often associated with severe and persistent behavioral problems, it is not a necessary characteristic for an adolescent to exhibit problem or delinquent behaviors. This interpretation is consistent with the conceptualization of psychopathy as distinct from antisocial behaviors in general due to the presence of maladaptive affective traits (Christian et al., 1997; Frick, 2003).

Callousness items were also expected to be particularly useful for the assessment of CU traits at the highest levels of the trait, but the low discriminating ability of Callousness items in general suggests that individuals across various levels of the underlying trait had relatively similar probabilities of endorsing these items. The lack of discriminating ability specifically at the highest levels of the CU construct was likely influenced by the overall low endorsement of CU traits in both samples. The Callousness subscale score was positively skewed ($skewness = .95$, $SE = .14$), indicating that the majority of participants endorsed low levels of callous traits within the observed range. The Uncaring and Unemotional subscale scores were more normally skewed [$skewness =$

.30 ($SE = .14$) and .09 ($SE = .14$), respectively], indicating a different approach to responses on the Callousness scale items.

A closer investigation of the items comprising these subscales may partially explain the relatively skewed Callousness scores. Most of the Callousness items are phrased in such a way that item endorsement requires acknowledgement and disclosure of socially undesirable characteristics (e.g., “I do not care if I get into trouble,” “I do not care who I hurt to get what I want”). In contrast, the Uncaring items are written as socially desirable characteristics which are later reverse-scored (e.g., “I feel bad or guilty when I do something wrong,” “I try not to hurt others’ feelings”). Thus, failure to endorse the Uncaring items does not directly indicate the presence of negative characteristics but instead suggests the absence of positive characteristics. The Uncaring items may be perceived as less threatening resulting in a greater willingness to respond honestly. Perhaps if greater levels of overall CU traits were endorsed, the Callousness items would better discriminate between different levels of the underlying trait and become particularly useful at the highest levels of CU traits.

Based on the poor discriminating ability of certain items, the overall information provided by the ICU in this sample was examined for exploratory purposes. The ICU provided the most reliable information at higher trait levels for the non-offender sample compared to the offender sample. In addition, the ICU provided more overall information for non-offenders. Therefore, among adolescents in a community sample, those reporting the highest overall CU traits (i.e., adolescents achieving the highest total ICU scores) appear to possess characteristics that are more closely related to the underlying CU construct compared to non-offending adolescents endorsing low or moderate trait levels.

Overall, the ICU provided more test information in the non-offender sample suggesting that the ICU estimates CU traits more reliably for non-offenders. Thus, the ICU appears to be more useful for measuring CU traits and more closely capturing the CU construct for non-offenders than offenders.

Differential Item Functioning

Both non-uniform and uniform DIF were observed in the current study. Some items did not discriminate in the same way across various levels of the underlying trait between groups (i.e., non-uniform DIF). Differences also existed between groups in the point along θ at which certain items became useful (i.e., uniform DIF). Eight items, representing all three subscales, exhibited either non-uniform or uniform DIF with one item (i.e., Item 4) exhibiting both types. This finding raises some issues for the use and interpretation of the ICU in different adolescent samples. ICU Callousness items 4 and 7 along with Unemotional item 22 demonstrated significant non-uniform DIF in such a way that suggests that these items were systematically more closely related to the underlying CU trait for non-offenders. Specifically, endorsement of these items by members of the non-offender group provides more useful information regarding CU traits than endorsement by the offender group. That is, a non-offender's endorsement of these items was more likely driven by the individual's actual CU trait level which corresponds to item difficulty.

Alternatively, there is less confidence that an offender's responses to those items reflected his or her level of CU traits. The offender may have responded in the same way to items 4, 7, and 22 despite having significantly lower or higher levels of CU traits compared to the non-offender. It is unclear why these items functioned differently

between groups given the item content which includes concern over hurting others, concern about being on time, and hiding emotions from others. Perhaps adolescent offenders interpreted these items differently based on their disposition at the time of data collection. For example, hiding one's feelings from others may be perceived as adaptive in a detention setting and not necessarily a truly psychopathy-linked characteristic. Though adolescent offenders demonstrated a similar likelihood of endorsing these items despite varying levels of CU traits within the offender sample, other ICU items with similar content (e.g., "I express my feelings openly") did not demonstrate non-uniform DIF between groups. This pattern may suggest that the observed differences between groups are a function of the wording or phrasing of the items and not necessarily differences in the underlying trait. The lack of significant non-uniform DIF across all items assessing similar content may also indicate that what appear to be congruent items are actually capturing distinct aspects of CU traits.

Regarding uniform DIF, endorsement of items 4, 9, 12, 13, 14, and 19 was less likely in the offender sample assuming overall trait levels equivalent to those of the non-offender sample. In other words, these items were "more difficult" for the offender sample compared to the non-offender sample. Consider the following hypothetical example using the construct of depression: A male and a female with identical levels of depression each complete a self-report measure of depressive symptoms. Despite having the same levels of depression, an assumption is that the female is more likely to endorse an item reading "I cry all the time" than the male because of possible gender differences in the manifestation of depressive symptoms. Therefore, this item is "more difficult" for depressed males than depressed females despite comparable levels of depression. In the

same way, these six ICU items were less likely to be endorsed (i.e., more difficult) by an adolescent offender than a non-offender with equivalent overall levels of CU traits.

The content of the six items exhibiting uniform DIF was investigated in hopes of determining why such differences existed between the two groups. These items represent all three ICU subscales (Callousness items 4, 9, and 12; Uncaring item 13; Unemotional items 14 and 19). Three of the items were reverse-scored. Item content involves lack of concern over getting in trouble or hurting others to achieve one's goal, ease of admitting to being wrong, and degree of emotional expressiveness. Using item 9 which assesses concern over getting in trouble as an example, results suggest that an adolescent offender was less likely to endorse caring about getting in trouble compared to a non-offender after controlling for CU trait level. One consideration when interpreting group differences in difficulty of Callousness items is the previously discussed skew of this subscale. However, skewness does not fully explain the differences between groups because low levels of callousness were endorsed in both samples, and uniform DIF was observed for Uncaring and Unemotional items as well. Another possible explanation takes into account the influence of environmental factors, specifically dispositional setting, on the adolescents' responses. Although both samples were informed that data collection was solely for research purposes, offenders may have been more reluctant to endorse high levels of CU traits for fear of perceived consequences. The offender sample consisted of males in a juvenile detention center and detained females awaiting trial. Compared to the community sample, offenders may have been more strongly motivated to present themselves favorably. An important finding of this study is that the eight items demonstrating DIF were not the most discriminating or difficult items. Therefore, these

items were not the best indicators for identifying individuals with varying levels of CU traits in either sample. The relatively poor item parameter estimates of these eight items combined with the observed differences in item functioning between groups sheds reasonable doubt on the specific usefulness of these items. Further research examining the functioning of these individual ICU items appears warranted.

Based on the pattern of uniform and non-uniform DIF among ICU items, it appears that DIF has an impact not only at the item level but also at the test level. DIF existed in such a way that seems to favor the use of the ICU in the non-offender sample over the offender sample. Higher underlying trait levels were necessary for certain items to be endorsed in the offender sample (i.e., greater difficulty of some items in the offender sample), whereas other items discriminated between levels of CU traits better in the non-offender sample. Therefore, one should exercise caution when comparing scores across groups based on the current study.

CU Traits as Assessed by the ICU

The ICU is the only known self-report measure designed to solely assess the affective features of psychopathy in adolescents. Other self-report measures, such as the APSD and the YPI, include items assessing both affective and behavioral dimensions of psychopathy. The presence of CU traits is believed to be central to the construct of psychopathy (Cleckley, 1976; Cooke & Michie, 1997; Hare, 1998), is associated with more severe and varied problem behaviors (Falkenbach et al., 2003; Frick, Cornell, et al., 2003), and is particularly useful for distinguishing between individuals with high levels of psychopathy-linked characteristics and those exhibiting antisocial behavior in general (Frick, 2003). Other IRT studies investigating psychopathy-linked traits in adolescents

have utilized item scores of the PCL:YV (Schrum & Salekin, 2006; Vincent, 2002) which are based on examiner ratings of items assessing affective, interpersonal, and behavioral features of psychopathy. Both Schrum and Salekin (2006) and Vincent (2002) concluded that the affective items were the most useful for discriminating between adolescents across varying levels of psychopathy. Thus, perhaps factors inherent to self-report assessment versus interview-based assessments contributed to the current results. For example, self-report assessment assumes adequate reading comprehension skills of the participants and that all participants interpret the items similarly. Also, interviewer ratings may be less likely to be influenced by social desirability, positive impression management, or defensiveness relative to adolescents' self-reports. Conversely, interviewer ratings may be subject to potential bias based on examinee-dependent variables such as behavioral problems, type of criminal offense, race, or gender, among others (Andershed, Hodgins, & Tengström, 2007; Heilbrun, Marczyk, DeMatteo, & Mack-Allen, 2007; Vitale & Newman, 2001).

Another important issue to consider is construct validity. Evaluating construct validity usually involves an examination of the measure's internal consistency and external validity (Cronbach & Meehl, 1955). Adequate reliability, or internal consistency, estimates for the ICU scores were obtained for both the offender and non-offender samples in the current study, providing modest support for the notion that the ICU assesses a unitary construct. There was no clear external criterion available to which the current sample's ICU scores could be compared to evaluate convergent or divergent validity. Involvement in the juvenile justice system could potentially be considered an external criterion based on previously established associations between CU traits and

delinquent behaviors. If one accepts this relation, adolescent offenders would be expected to produce higher mean ICU scores than non-offenders which was not the case in the current study. Future research examining the utility of the ICU and its individual items is needed to adequately assess construct validity. Comparing ICU scores to analogous scores from well-established psychopathy measures (e.g., PCL: YV, YPI) is one way that construct validity should be investigated.

Practical Implications for the Assessment of CU Traits in Adolescents

As the construct of psychopathy in adolescence gains empirical ground, the need for reliable, valid, and efficient assessment methods becomes increasingly necessary. Screening instruments, as opposed to interview-based assessments, require little training for the examiner and are significantly less time-consuming for examiners and participants. Self-report measures offer the additional benefit of assessing certain characteristics which cannot be readily observed by an outside examiner (e.g., callousness, uncaring). Based on the current results, ICU items 2 and 10 (i.e., “What I think is ‘right’ and ‘wrong’ is different from what other people think” and “I do not let my feelings control me,” respectively) do not appear to be related to overall CU traits. This finding is consistent with previous research examining the psychometric properties of the ICU (Kimonis, Frick, Skeem et al., 2008). Therefore, omission or modification of these two items appears warranted. Of the remaining items, several demonstrated adequate discriminating ability across a relatively wide range of the underlying trait suggesting that most of the ICU items are appropriate for a screening instrument. However, interpretations of ICU scores beyond the purposes of a screening measure should be made cautiously. Among adolescents with equivalent trait levels, certain items

may be more difficult and provide less information regarding CU traits in an adolescent offender sample compared to a non-offender or community sample. Therefore, an adolescent from a non-offender sample could produce ICU scores greater than adolescent offenders despite actually having lower underlying levels of CU traits, which could lead a practitioner to erroneously interpret ICU scores as clearly indicative of elevated levels of maladaptive traits or predictive of a greater likelihood of future problem behaviors. Conversely, if an adolescent offender's total score is lower than an expected average, or norm, found in the community, examiners may mistakenly conclude that further evaluation of psychopathy-linked traits is not warranted.

Furthermore, the current exploratory factor analysis results indicate that ICU items may be loading onto separate factors, or subscales, based on the item-format rather than shared item content or conceptually distinct aspects of CU traits. The combination of reverse-scored and regular-scored items on the ICU may be distorting the factor structure resulting in misleading delineations of individual components of CU traits. If the current reverse-scored ICU items were rewritten to be scored in the same direction as the remaining items, factor analyses may yield different ICU subscales and potentially improve some of the psychometric properties of the ICU. Despite these issues and in light of the current ICU subscales, it appears that the ICU is appropriate for identifying CU traits in adolescents but particularly useful for those traits conceptualized as uncaring (e.g., lack of concern about academic/work performance, not trying to avoid hurting others' feelings).

Limitations and Directions for Future Research

One limitation of this study is the use of a relatively modest sample size to conduct IRT analyses. There is no clear rule for determining the sample size needed to estimate item parameters. Some researchers suggest that the graded response model can be estimated with as few as 250 participants (Embretson & Reise, 2000). One recommendation for obtaining IRT estimates for polytomous items is to ensure that participant responses represent every response category for all items (Reeve & Fayers, 2005), which was achieved in the current sample. Second, the present sample consisted solely of adolescents from the southern United States which reduces generalizability to adolescents in other geographical regions. Also, the conceptualization of participants as non-offenders was based on their high school enrollment status at the time of data collection. It is possible that some of these individuals had prior arrests as well as had engaged in illegal activities for which they were not arrested nor charged. However, it is reasonable to assume that illegal or delinquent acts would be present in any adolescent sample and that the behaviors of adolescents placed in a juvenile justice setting (e.g., offender sample) represent more severe or frequent delinquency than those of community adolescents.

Regarding the CFA results, it is possible that demographic variables, including cultural and language differences, contributed to the weak fit of the previously established model proposed by Essau and colleagues (2006) and subsequently replicated in a different sample (Kimonis, Frick, Skeem et al., 2008) to the current sample data. Additional research regarding the factor structure of the ICU in other North American adolescent samples is necessary. In addition, the sample data used in the current study did

not fully meet the assumptions of IRT, namely unidimensionality and local independence. However, the obtained item parameter estimates are still considered interpretable. Reeve and Fayers (2005) noted that IRT models are generally resistant, or robust, to the effects of assumption violations on parameter estimates.

The current results highlight the need for further research surrounding the assessment and conceptualization of CU traits in adolescents. Future research could address limitations of the current study by investigating the utility of the ICU with a larger sample and in various settings and populations (e.g., correctional/juvenile justice settings, diverse geographical regions). It would be interesting to differentiate between adolescents with violent or aggressive offenses versus non-violent offenders in terms of item functioning. Future research examining the utility of ICU items and potential DIF depending on gender, race, socioeconomic status, and other demographic variables is necessary for gaining a fuller understanding of the manifestation of CU traits and possible idiosyncrasies of self-report assessment of these traits in various adolescent populations. Items exhibiting greater difficulty and discriminating abilities, along with those demonstrating DIF, seem especially worthy of further examination in other populations.

Conclusion

In general, the current findings support the use of the ICU as a self-report measure of CU traits in adolescents. Of particular interest are the observed differences in discriminating ability and item difficulty across ICU items as well as the differences in item functioning between two groups of adolescents. In light of the growing interest in adolescent psychopathy among researchers and practitioners, a deeper understanding of

the presentation of psychopathy-linked traits in adolescents as well as reliable and valid measurement of these traits is increasingly necessary. Investigations of the functioning of self-report measures of psychopathy-linked traits in adolescents, particularly individual items comprising these measures, not only enhance theoretical conceptualizations of psychopathy but can also lead to methodological improvements regarding the way in which these traits are operationally defined and measured as well as in how they ultimately inform intervention decisions.

FOOTNOTE

¹Researchers interested in investigating the extent to which random responding, or guessing, has influenced an individual's responses across items can include a third parameter in the IRT model. The c parameter refers to the *pseudo-guessing* parameter (Zickar & Broadfoot, 2009). However, IRT researchers suggest that the c parameter is likely to be more useful when conducting IRT analyses on tests measuring an ability level rather than instruments designed to assess personality traits such as the measure of psychopathy-linked characteristics of interest in the present study.

APPENDIX

ITEMS AND SUBSCALES OF THE INVENTORY OF CALLOUS-UNEMOTIONAL
TRAITS (ICU; FRICK, 2003)Callousness Items

- 2. What I think is “right” and “wrong” is different from what other people think.
- 4. I do not care who I hurt to get what I want.
- 7. I do not care about being on time.
- 8. I am concerned about the feelings of others.*
- 9. I do not care if I get into trouble.
- 10. I do not let my feelings control me.
- 11. I do not care about doing things well.
- 12. I seem very cold and uncaring to others.
- 18. I do not feel remorseful when I do something wrong.
- 20. I do not like to put the time into doing things well.
- 21. The feelings of others are unimportant to me.

Uncaring Items

- 3. I care about how well I do at school or work.*
- 5. I feel bad or guilty when I do something wrong.*
- 13. I easily admit to being wrong.*
- 15. I always try my best.*
- 16. I apologize (“say I am sorry”) to persons I hurt.*
- 17. I try not to hurt others’ feelings.*
- 23. I work hard on everything I do.*
- 24. I do things to make others feel good.*

Unemotional Items

- 1. I express my feelings openly.*
- 6. I do not show my emotions to others.
- 14. It is easy for others to tell how I am feeling.*
- 19. I am very expressive and emotional.*
- 22. I hide my feelings from others.

* = Reverse-scored items

REFERENCES

- Ansel, L. (2009). *An analysis of four self-report measures of adolescent psychopathy: Exploring unique prediction of delinquency and aggression* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg.
- Andershed, H. A., Hodgins, S., & Tengström, A. (2007). Convergent validity of the Youth Psychopathic Traits Inventory (YPI): Association with the Psychopathy Checklist: Youth Version (PCL:YV). *Assessment, 14*, 144-154.
- Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: Initial test of a new assessment tool. In E. Blaauw & L. Sheridan (Eds.), *Psychopaths: Current international perspectives* (pp. 131-158). The Hague: Elsevier.
- Arbuckle, J. (2010). IBM SPSS Amos (Version 19) [Computer software]. Meadville, PA: Amos Development Corporation.
- Bolt, D. M., Hare, R. D., Vitale, J. E., & Newman, J. P. (2004). A multigroup item response theory analysis of the Psychopathy Checklist—Revised. *Psychological Assessment, 16*, 155-168.
- Brown, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (p. 136-162). Newbury Park, CA: Sage.
- Christian, R., Frick, P. J., Hill, N., Tyler, L. A., & Frazer, D. (1997). Psychopathy and conduct problems in children: II. Subtyping children with conduct problems based on their interpersonal and affective style. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 233-241.

- Cleckley, H. (1976). *The mask of sanity*. (5th ed.). St. Louis, MO: Mosby.
- Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist—Revised. *Psychological Assessment*, 9, 3-14.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Douglas, K. S., Vincent, G. M., & Edens, J. F. (2006). Risk for criminal recidivism: The role of psychopathy. In C. Patrick (Ed.), *Handbook of psychopathy* (pp. 533-554). New York, NY: Guilford.
- Edens, J. F., Skeem, J. L., Cruise, K. R., & Cauffman, E. (2001). Assessment of “juvenile psychopathy” and its association with violence: A critical review. *Behavioral Sciences and the Law*, 19, 53-80.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, 13, 454-469.
- Falkenbach, D. M., Poythress, N. G., & Heide, K. M. (2003). Psychopathic features in a juvenile diversion population: Reliability and predictive validity of two self-report measures. *Behavioral Sciences and the Law*, 21, 787-805.
- Forth, A. E., Brown, S. L., Hart, S. D., & Hare, R. D. (1996). The assessment of psychopathy in male and female noncriminals: Reliability and validity. *Personality and Individual Differences*, 20, 531-543.
- Forth, A. E., Hart, S. D., & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment*, 2, 342-344.

- Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Hare Psychopathy Checklist: Youth Version*. Toronto, Ontario, Canada: Multi-Health Systems.
- Frick, P. J. (2003). *The Inventory of Callous-Unemotional Traits*. Unpublished rating scale, The University of New Orleans.
- Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology*, 31, 457-470.
- Frick, P. J., & Hare, R. D. (2001). *The antisocial process screening device (APSD)*. Toronto: Multi-Health Systems.
- Hambleton, R. K., Swaminathan, H., Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, 1, 111-119.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist—Revised*. Toronto, Ontario: Multi-Health Systems.
- Hare, R. D. (1998). Psychopathy, affect and behavior. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 105-137). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Heilbrun, K., Marczyk, G., DeMatteo, D., & Mack-Allen, J. (2007). A Principles-based approach to forensic mental health assessment: Utility and update. In Goldstein, A. M. (Ed.), *Forensic psychology: Emerging topics and expanding roles* (pp. 45-72). Hoboken, NJ: John Wiley & Sons, Inc.

- Jones, S., Cauffman, E., Miller, J. D., & Mulvey, E. (2006). Investigating different factor structures of the Psychopathy Checklist: Youth Version: Confirmatory factor analytic findings. *Psychological Assessment, 18*, 33-48.
- Kimonis, E. R., Frick, P. J., Munoz, L. C., & Aucoin, K. J. (2008). Callous-unemotional traits and the emotional processing of distress cues in detained boys: Testing the moderating role of aggression, exposure to community violence, and histories of abuse. *Development and Psychopathology, 20*, 569-589.
- Kimonis, E. R., Frick, P. J., Skeem, J. L., Cruise, K., Aucoin, K. J., Morris, A. S., Munoz, L. C., & Marsee, M. A. (2008). Assessing callous-unemotional traits in adolescent offenders: Validation of the inventory of callous-unemotional traits. *International Journal of Law and Psychiatry, 31*, 241-252.
- Kosson, D. S., Cyterski, T. D., Steuerwald, B., Neumann, C. S., & Walker-Matthews, S. (2002). The reliability and validity of the Psychopathy Checklist: Youth Version (PCL:YV) in nonincarcerated adolescent males. *Psychological Assessment, 14*, 97-109.
- Loney, B. R., Taylor, J., Butler, M. A., & Iacono, W. G. (2007). Adolescent psychopathy features: 6-year temporal stability and the prediction of externalizing symptoms during the transition to adulthood. *Aggressive Behavior, 33*, 242-252.
- Lynam, D. R. (1997). Pursuing the psychopath: Capturing the fledgling psychopath in a nomological net. *Journal of Abnormal Psychology, 106*, 425-438.
- Lynam, D. R., Caspi, A., Moffitt, T. E., Loeber, R., & Stouthamer-Loeber, M. (2007). Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of Abnormal Psychology, 116*, 155-163.

- Marsee, M. A. (2008). Reactive aggression and posttraumatic stress in adolescents affected by Hurricane Katrina. *Journal of Clinical Child & Adolescent Psychology, 37*, 519–529.
- Marsee, M. A., & Frick, P.J. (2007). Exploring the cognitive and emotional correlates to proactive and reactive aggression in a sample of detained girls. *Journal of Abnormal Child Psychology, 35*, 969-981.
- McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24*, 99-114.
- Moffitt, T. E. (2003). Life-course-persistent and adolescent-limited antisocial behavior: A 10-year research review and research agenda. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 49-75). New York, NY: Guilford Press.
- Muthén, L.K., & Muthén, B.O. (2006). Mplus [Computer software]. Los Angeles, CA: Muthén and Muthén.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- Poythress, N., Dembo, R., Wareham, J., & Greenbaum, P. (2006). Construct validity of the Youth Psychopathic Traits Inventory (YPI) and the Antisocial Process Screening Device (APSD) with justice-involved youths. *Criminal Justice and Behavior, 33*, 26-55.
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire items and scale properties. In P. M. Fayers & R. D. Hays

- (Eds.), *Assessing quality of life in clinical trials: Methods and practice* (2nd ed., pp. 55-73). New York, NY: Oxford University Press.
- Salekin, R. T. (2008). Psychopathy and recidivism from mid-adolescence to young adulthood: Cumulating legal problems and limiting life opportunities. *Journal of Abnormal Psychology, 117*, 386-395.
- Salekin, R. T., Rogers, R., & Sewell, K. (1996). A review and meta-analysis of the psychopathy checklist and psychopathy checklist-revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*, 203-215.
- Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monographs, 34* (4, Pt. 2, Whole No. 17).
- Schrum, C. L., & Salekin, R. T. (2006). Psychopathy in adolescent female offenders: An item response theory analysis of the Psychopathy Checklist: Youth Version. *Behavioral Sciences and the Law, 24*, 39-63.
- Schumacker, R., & Lomax, R. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). New York, NY: Taylor & Francis Group.
- Thissen, D. (1991). Mutlilog (Version 6) [Computer program]. Morresville, IN: Scientific Software.
- Thissen, D. (2001). IRTLRFIF: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning (Version 2) [Computer software]. LL Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill.

- Thissen, D., Chen, W-H., & Bock, R. D. (2003). MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory (Version 7) [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.
- Vincent, G. M. (2002). *Investigating the legitimacy of adolescent psychopathy assessments: Contributions of item response theory* (Unpublished doctoral dissertation). British Columbia, Canada: Simon Fraser University.
- Vitale, J. E., & Newman, J. P. (2001). Using the Psychopathy Checklist—Revised with female samples: Reliability, validity, and implications for clinical utility. *Clinical Psychology: Science and Practice*, 8, 117-132.
- Vitale, J. E., Smith, S. S., Brinkley, C. A., & Newman, J. P. (2002). The reliability and validity of the Psychopathy Checklist—Revised in a sample of female offenders. *Criminal Justice and Behavior*, 29, 202-231.
- Zickar, M. J., & Broadfoot, A. A. (2009). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York, NY: Routledge/Taylor & Francis Group.